

Human genome projects: work in progress

Whatever doubts there may be about the timing, public celebrations of genome projects are well deserved. Researchers should now be left to complete significant tasks that remain, and to tidy up troublesome errors in their databases.

Francis Collins and his colleagues at the head of the publicly funded Human Genome Project deserve praise, not only for their science, but for their openness, their discipline, their international perspective and for their systematic approach. That surely represents a model to be supported for future large programmes in biology. Craig Venter and his colleagues at the private company Celera deserve credit for the technical imagination and competitiveness without which the world would probably be waiting a lot longer for these results to emerge. And where, as here, competitiveness and technical complementarity go hand in hand, science benefits too.

But, given the significantly incomplete nature of the research in both projects, why the presidential and prime ministerial fanfares (see page 983), and why this week in particular?

Anyone interested in the human genome projects will have become familiar with two sorts of public announcements: those in scientific journals and those not linked to any peer-reviewed publication. This week's extravagant examples of the latter reached an all-time zenith or nadir, according to taste. One might have been forgiven for forgetting that there are months to go before even a draft sequence will be scientifically useful. But that fact need not prevent congratulations going where they are already due.

Both private and public projects have issued a series of announcements that have helped maintain a high public profile, with one eye on share prices, no doubt, in the former case, and on political representatives in the latter. None has been as scientifically arbitrary in its timing as this week's festivities, where even the scientifically useful 'draft' is

still incomplete by previously suggested definitions. But there has been increasing pressure on both projects to deliver something that their private or public investors — not to mention the media — could celebrate. And, as the public are represented by politicians only too pleased to identify themselves with the achievement, in the end there was no way the projects could wait for the months required to make the draft publishable. Above all, the bickering needed to be stopped.

Even sceptical purists must hope that this week's celebration will not only boost public confidence to appropriate levels but will also take the spotlight off the two, in order that completion of a scientifically useful draft can be most effectively achieved.

But even that can be expected to leave significant gaps. Not only will that sequence be incomplete (by definition), but the published 'annotation', the identification of genes and of their function, will contain many uncertainties (see page 984). If, despite Celera's commercial interests and status, it and the public project can find a way of benefiting from each other so as to hasten completion of the sequence and to help speed up its public annotation, so much the better.

But there remains a related problem. More attention must be given to the accuracy of public databases into which sequences and annotation are being deposited. These contain errors that could themselves take years to clean up. Biologists encountering such errors should be doing more to provide the feedback to databases and to original depositing authors to achieve corrections. Better still, Bill Clinton, Tony Blair and others should press for dedicated funds for that unglamorous but essential curatorial purpose. ■

Bad clinical practices

When it comes to research, Germany's medical profession may have a lot to answer for.

Is there just one rotten apple in the barrel? Or is the barrel itself contaminated? There are many in Germany who believe that the suspicions of scientific misconduct raised against three clinical researchers, originally from Freiburg (see *Nature* **405**, 871–872; 2000), is a sign that the system of clinical research and promotion in Germany provides a breeding ground for bad scientific practice.

Suspensions against the researchers were revealed during a two-year independent analysis of research papers stemming from the laboratory of prominent Freiburg clinician Roland Mertelsmann, in the wake of the scandal surrounding Mertelsmann's former right-hand man, Friedhelm Herrmann. A couple of years ago, Herrmann's meteoric career crashed following charges that data had been fabricated in many of his research papers. The careers of two others which were also launched from the Mertelsmann group are now in question after the task force that conducted the analysis presented evidence of data manipulation in their *Habilitation* theses. (*Habilitation* is a high-level research qualification required

for an academic career, particularly in the clinical sciences.)

Charges of scientific negligence have been brought against Mertelsmann himself. Whatever the outcome, there is a glaring paradox in the clinical system in Germany: although a long publication list is required for any top clinical position — even for those positions where no research will ever be done — there is no formal requirement for any training in research methods. On top of this, clinicians are not given leave from their wards for their research; data must be collected after hours. That is a blueprint for trouble.

The solution is as obvious as the paradox: *Habilitation* should be scrapped. Promotion should be based on an appropriate assessment of quality, not quantity, of research, balanced against bedside skills. And a PhD or MD/PhD programme for clinicians should be introduced. Despite public censure, the Bavarian Chamber of Physicians allows Herrmann to continue practising as a clinician in Munich. So why are there no positive moves to change? Because the medical profession is deeply conservative, protective and arrogant. ■



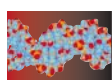
Genome special

After fourteen years of work, the human genome sequence has now reached draft form. As Celera and the Human Genome Project declare a ceasefire, attention is turning to what to do with the data p983

Red liquid
Possible water
courses on Mars
galvanize NASA
p987

Patent saved
Alzheimer's
research using mice
protected in court
p989

World leaders heap praise on human genome landmark



Colin Macilwain, Washington

It was portrayed as the scientific rivalry to end all scientific rivalries. But after two years of often acrimonious competition, the principal players in the Human Genome Project (HGP) came together on Monday this week with Craig Venter, president of Celera Genomics of Rockville, Maryland, to praise one another's achievements — and to bask in the adulation of their political leaders.

Those achievements — the compilation of a 'working draft' of the human genome by the HGP and Celera's 'first assembly' of a complete genome sequence — may in reality represent arbitrary milestones on the way to the goal of fully deciphering the human genetic code. But for those scientists who have devoted their careers to genomics since the launch of the HGP more than a decade ago, a public celebration is welcome.

In Washington, President Bill Clinton hosted an event at the White House with Venter and Francis Collins, director of the National Human Genome Research Institute in Bethesda, Maryland — joined by video link from London by Prime Minister Tony Blair. In Paris, Tokyo and Bonn, other members of the international HGP consortium made their own announcements. The Wellcome Trust, the main backer of the Sanger Centre near Cambridge, responsible for one third of the HGP sequence, hosted its own press conference in London.

Turning to James Watson, discoverer with Francis Crick of the structure of DNA and head of the HGP until 1992, Clinton described the 1953 *Nature* paper, which noted the structure's "novel features, which are of considerable biological interest" as "one of the great understatements of all time". This time, understatement was in short supply. "With this profound new knowledge, humankind is on the verge of gaining immense, new power," said Clinton.

Mike Dexter, director of the Wellcome Trust, went so far as to describe the project's significance as surpassing the invention of the wheel. "I can see technology making the wheel obsolete," he said. "But this code will be useful and used as long as humans exist."



Truce: Venter (left) and Collins bury their differences for the joint announcement of the draft map.

For some scientists, this may seem like hyperbole — particularly as the value of both the HGP and Celera drafts, in their current form, remains unclear. The sequences are difficult to compare, because of the two groups' different approaches. Celera's 'whole-genome shotgun' strategy relied on shattering the entire genome into tiny pieces, sequencing these *en masse*, and then using sophisticated computer algorithms to put the pieces back together. The HGP cloned

larger overlapping fragments of the genome into some 24,000 bacterial artificial chromosomes, and sequenced these one by one.

The HGP claims to have cloned 97% of the genome, and sequenced 85% of it to draft standard. That falls short of the 90% figure set as a target for the working draft — although this should be achieved within a few weeks. On average, the HGP has sequenced each base seven times over (7X coverage). Celera's assembly gives 4.6X coverage — some way

The story so far...

March 1986 US Department of Energy (DoE) holds meeting in Santa Fe to discuss plans to sequence the human genome.

April 1987 Advisory panel at DoE suggests spending \$1 billion on genomics over seven years. Small-scale genome programme starts within DoE.

February 1988 Support for Human

Genome Project (HGP) appears in report from US National Research Council.

March 1988 National Institutes of Health (NIH) announces first genome sequencing efforts.

April 1988 Congressional Office of Technology Assessment endorses Human Genome Project.

September 1988 Office of Human Genome Research established in NIH. Nobel prizewinner James Watson, co-discoverer of the double helix structure of DNA, appointed as its head.

October 1989 NIH establishes National Center for Human Genome Research (NCHGR), again with Watson at its helm.

April 1990 NIH and DoE publish five-year mapping and sequencing plan for 1990–1995 costing a projected \$200 million a year.

July 1991 Craig Venter, then at the National Institute of Neurological Disorders and Stroke, part of NIH, reveals that NIH has applied for patents on isolated DNA fragments from

brain tissue, sequenced in his lab. Watson states his opposition to these 'expressed sequence tag' patents, saying that automated sequencing machines "could be run by monkeys".

April 1992 Watson resigns as head of NCHGR in the wake of this dispute, and following allegations that his holdings of

from the 10X it originally promised in 1998, although the company says its sequence covers 99% of the genome. On the basis of the data presented publicly, it is impossible to verify whether Celera's assembly is correctly oriented and ordered throughout the genome. But Celera has also produced a second map by incorporating data from the public project — which will increase its depth of coverage and allow it to check its shotgun assembly.

Despite the preliminary nature of both sets of data, the White House has been encouraging the two projects to bury their differences and declare their drafts complete. US politicians were appalled at the media portrayal of the sequencing project as a battle between Celera and the HGP. Clinton's aides hope that a joint announcement will end the rancour and lead to greater public recognition of the achievements of both projects.

The rapprochement between Collins and Venter was brokered by senior figures including Eric Lander, director of the Whitehead Institute at the Massachusetts Institute of Technology, one of the main sequencing centres for the HGP, and Ari Patrinos, head of biological and environmental research at the Department of Energy, who hosted meetings over beer and pizza at his home in Rockville. The agreement has four parts: Monday's choreographed joint announcement; a pledge

to publish the two draft sequences, simultaneously but separately, later in the year; a loosely defined plan to hold a joint meeting of the two research teams after publication; and a promise to keep open lines of communication between the HGP and Celera.

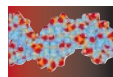
At the White House event, Collins struck a spiritual note: "It is humbling for me and awe-inspiring to realize that we have caught the first glimpse of our own instruction book, previously known only to God." Venter was philosophical: "The complexities and wonder of how the inanimate chemicals that are our genetic code give rise to the imponderables of

the human spirit should keep poets and philosophers inspired for the millenniums."

For scientists working on the HGP, the main hope is that the task of finishing the genome sequence does not get subordinated to other activities. Parallels drawn with the Apollo lunar programme — which soon fizzled out after the space race was won — provide a warning. "What we most want to avoid is the fate of achieving this heroic goal at such great cost and to the neglect of the long-term goals," says Maynard Olson, a geneticist at the University of Washington in Seattle. ■

Additional reporting from David Dickson in London.

Draft data leave geneticists with a mountain still to climb



Declan Butler and Paul Smaglik

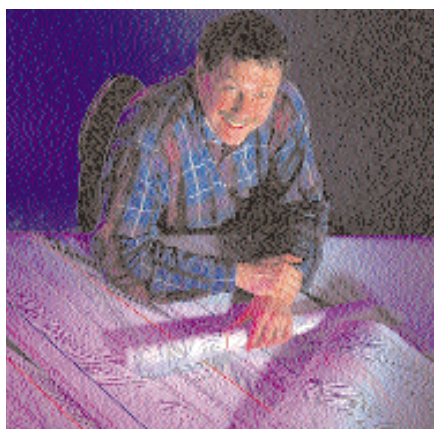
Now the race to obtain a draft sequence of the human genome has been declared an honourable draw, attention will switch to the task of finishing the sequence and 'annotating' the entire genome — characterizing all its genes and working out their functions. The annotation is so formidable that it may need the largest Internet 'collaboratory' yet attempted.

Given that Celera has now stopped sequencing, the task of finishing the genome — in which, to ensure accuracy, each base has been sequenced 10 times over (10X coverage) — will fall to the public Human Genome Project (HGP). In that regard, says Tim Hubbard of the Sanger Centre at Hinxton, near Cambridge, the HGP got a pleasant surprise last weekend, when its data were subjected to a "brute force" computer analysis. Hubbard had expected to find that the HGP had sequenced the genome to an average depth of 5X, but instead, a figure of 7X emerged. This, and the fact that the draft seems to contain fewer gaps than expected, bodes well for finishing the genome ahead of

the stated 2003 deadline, says Hubbard.

But annotation poses a much bigger challenge. The first step is to identify all of the protein-coding regions, which will give a good idea of how many genes there are. Most geneticists think the figure lies somewhere between 35,000 and 150,000. Beyond that will come detailed studies of the structure of individual genes, including their regulatory elements, and attempts to assign functions to them.

David Lipman, director of the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland, believes that the draft sequence will allow researchers to use computational tools to pinpoint the position of many of the gene fragments catalogued in cDNA libraries of expressed genes. In many cases, it will then be possible to extract an entire gene from the draft sequence — and by comparison with other genes, begin to establish its function. But many biologists are unconvinced. "The current perception is that annotating finished sequence is much less difficult than annotating 'sequence in progress,'" says Richard Gibbs of Baylor College of Medicine in Houston. "And no matter how



Lander: helped to broker a genomic ceasefire.

SAM OGDEN

biotech stocks represented a conflict of interest. Francis Collins of the University of Michigan appointed as his replacement.

June 1992 Venter leaves NIH to set up The Institute for Genomic Research (TIGR) in Rockville, Maryland. SmithKline Beecham provides \$125 million to

finance TIGR and to develop its findings commercially through a company called Human Genome Sciences.

July 1992 Britain's Wellcome Trust emerges as a major player in genomics by announcing funding of £50 million for projects including sequencing of the nematode worm

Caenorhabditis elegans. HGP is by this time a global endeavour, involving government- and charity-funded scientists from many developed nations.

October 1993 NIH and DoE publish revised plan for 1993–1998. Goal set of 80 megabases of DNA sequence by end of 1998. Full completion of human

genome sequence set for 2005.

October 1993 Wellcome Trust and UK Medical Research Council open Sanger Centre at Hinxton Hall, south of Cambridge, to sequence the human genome and those of model organisms.

September 1994 French and

American researchers publish a complete genetic linkage map of the human genome, one year ahead of schedule.

December 1995 Another collaboration, again led by scientists from the United States and France, publishes a physical map of the human genome containing

15,000 marker sequences.

February 1996 At a meeting in Bermuda, international HGP partners agree to release sequence data into public databases within 24 hours.

April 1996 International consortium announces complete

genome sequence of the yeast *Saccharomyces cerevisiae*.

January 1997 NCHGR renamed as National Human Genome Research Institute.

June 1997 TIGR breaks links with Human Genome Sciences following tensions over publication policy.

May 1998 Venter announces formation of a company — later named as Celera — to sequence human genome "within three years". Venter says he will use an ambitious 'whole-genome shotgun' method, but Celera's data access policy will not follow the Bermuda declaration.



Carry on sequencing: genome scientists, here preparing DNA for analysis, will get no respite.

you cut it, the draft is sequence in progress."

Even with the finished sequence in hand, experience with the two human chromosomes for which this has been achieved — numbers 21 and 22 — indicates that annotating the genome will be a mammoth task. "With 21 and 22 it was not possible to reliably identify and delineate all of the genes," says Philip Green, a biocomputing expert at the University of Washington in Seattle.

In the case of the genome of the fruitfly *Drosophila*, annotation was kickstarted by a two-week 'jamboree' held at Celera. This brought together over 40 academic fly geneticists and 50 Celera scientists, and compared the outcome of dozens of different annotation techniques. This experience should serve Celera well. "We basically trained their annotation team to annotate the human genome," observes Martin Reese, formerly of the *Drosophila* Genome Center at the Lawrence Berkeley National Laboratory in California and now with ValiGen, a company near Paris.

The news that Celera and HGP researchers will hold a joint scientific meeting after publishing simultaneous papers of their

Work in progress: data have accumulated rapidly but the public sequence is far from finished.

draft sequences (see lead story) initially raised hopes of a similar human jamboree. However, as HGP head Francis Collins pointed out to *Nature*, Celera cannot really share its annotation, as it will be its core product for sale to its subscribers. Rather, the meeting is expected to look at discrepancies between the public and private sequences with the goal of 'cleaning up' one another's data.

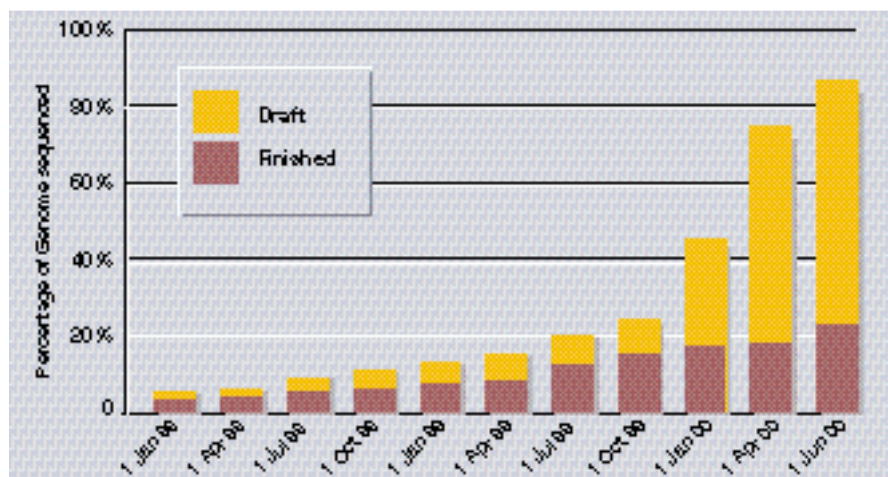
Celera has said little publicly about its annotation capacity, but it uses specialized software to combine the output of multiple gene finding tools — mostly those available to the public sector. But while Celera's annotation team is at the cutting edge, many experts argue that no single team is currently in a position to annotate the entire genome. "No one really knows how to do it completely," says John Quackenbush of The Institute for Genomic Research in Rockville, Maryland.

On the public side, annotating the genome might mean a rethink on how the HGP's data are organized. Lipman acknowledges that the main sequence database, NCBI's Genbank, has its limitations. "It does not represent what we know of biology at any given time," he says. "It only represents what the author put in." Indeed, while scientists deposit data in Genbank because many journals make this a condition for publication, some do not bother to correct and update it.

"With annotation we will need much more active curation," says Lipman. Many experts believe this may require a 'collabora-

tory' approach, using the Internet to leverage the talent of biologists worldwide. The NCBI intends to set up a system in which named biologists around the world will 'adopt' a gene or gene family, becoming the curators responsible for gathering information from the wider research community. But Lipman remains against the idea of a free-for-all in which any biologist can annotate the genome — the problem, he says, is that most do not fully understand database syntax, and so tend to make errors when they input data. "What we really want is their knowledge," says Lipman.

The Ensembl annotation project, run by the Sanger Centre and the European Bioinformatics Institute, is plotting a genuinely distributed effort. Hubbard foresees a system where a geneticist in Germany could annotate a gene online, and have his or her interpretation challenged almost in real time by a biologist in Boston. Ensembl's vision has been inspired by a radical suggestion, made by Tom Sleazak of the Lawrence Livermore National Laboratory in California and Lincoln Stein of the Cold Spring Harbor Laboratory on Long Island, to use 'Napster' technology for genome annotation. This allows computer users worldwide to share MP3 music files, and could, in theory, let biologists share and annotate genome data (see *Nature* 404, 694; 2000). If these ideas catch on, the genome project's future could be one of annotation by anarchy. ■



SOURCE: EBI/NCBI

May 1998 Wellcome Trust responds by announcing that it will double its support for HGP, taking on responsibility for one third of the sequencing.

October 1998 NIH and DoE publish goals for 1998–2003: one third (1 gigabase) of human sequence

and 'working draft' of the remainder of the genome by end of 2003, full sequence by end of 2003.

December 1998 Researchers in Britain and the United States announce genome sequence of *C. elegans*.

December 1998 NIH and Wellcome

Trust block proposed collaboration between Celera and DoE, arguing that the terms would conflict with HGP's open data access policy.

March 1999 NIH brings forward planned date for working draft to spring 2000. NIH and Wellcome Trust announce

end of 'pilot phase' and start of full sequencing. Costs reduced to 20–30 cents per base.

September 1999 NIH launches project to sequence mouse genome.

November 1999 HGP celebrates sequencing of one-billionth base of human DNA.

December 1999 First complete human chromosome sequence — for chromosome number 22 — published. Celera and HGP discuss possible collaboration.

January 2000 Celera announces compilation of DNA sequence covering 90% of human genome.

March 2000 Celera and academic collaborators release "substantially complete" sequence of the fruitfly *Drosophila melanogaster*, achieved using whole-genome shotgun method.

March 2000 Plans for HGP and Celera to collaborate founder amid

considerable acrimony. Data access policy is again the stumbling block.

March 2000 HGP announces successful sequencing of two billion bases of human genome.

April 2000 Celera announces completion of 'raw sequencing stage' of

human genome from one individual.

May 2000 HGP consortium led by German and Japanese researchers publish complete sequence of chromosome 21.

June 2000 HGP and Celera jointly announce working draft of human genome sequence. **D. D.**

Call for monitoring plan on German GM crops

Alison Abbott & Ulrike Hellerer, Munich

An impasse between government ministers over the cultivation of genetically modified (GM) crops has led the German chancellor, Gerhard Schröder, to intervene. Last week, he asked companies to agree voluntarily to grow their licensed crops only within the framework of a three-year government-based research and monitoring programme.

The move came despite unresolved differences within the coalition government over whether such a programme should be carried out — and, if so, how it should be run. Talks between the Social Democrat-led ministries of research and agriculture, both of which generally favour the introduction of commercial agricultural biotechnology, and the Green-led ministries of health and environment, which generally oppose it, are expected to reach agreement on these issues this week.

Speaking at the opening of an exhibition on plant breeding, organized as part of the EXPO 2000 world fair in Hannover, Schröder invited agricultural companies holding, or hoping to hold, licences to grow GM crops to negotiate with the government. These discussions would cover how such a programme should be run, what research should be carried out, which environmental impacts should be monitored, and who should pay.

Confirming that research into the applications of genetic engineering in both health and agriculture remains a government priority, Schröder also said that consumers need to be persuaded of the safety of GM foods, and not forced to accept them.

No details have yet been provided of how an experimental programme might proceed. But according to a spokesman from the Robert Koch Institute (RKI), which licenses field trials for the health ministry, consumer groups, as well as representatives from the non-commercial scientific environment, could eventually become involved. He predicts that the programme will be in place before the start of next year's growing season.

Response from industry, which has been left in a position of great uncertainty after the government blocked the first German commercial licence — for Novartis's insect-resistant *Bt* maize, last February (see *Nature* 403, 821; 2000) — has been positive. Novartis says that it would "expect to participate" in a research programme, and KWS, a major seed manufacturer, said that the programme would provide reassurance to all stakeholders, including industry, environmentalists and consumers.

The RKI itself is keen to participate in the government-level programme. It has long and vociferously complained about its advice being ignored by the health ministry, despite its legal status as the ministry's official scientific advisor. It had, for example, advised the



Green reaper: protesters last week dumped GM rapeseed at the German environment ministry.

ministry that there were no scientific grounds to block Novartis's commercial licence.

The new monitoring programme would bear some similarities to the farm-scale GM plant trials being run jointly by industry and the Department of the Environment, Transport and the Regions in the United Kingdom. But Schröder's plan, according to the RKI, will not necessarily imply a moratorium on

commercial cultivation. Instead it will be part of "a slow and careful introduction of farm-scale production". Nor will it necessarily involve destruction of the crops after the test period, as happens in the United Kingdom. Instead, the crops might be used to test processing techniques.

But environmentalists are far from happy. Stefan Flthmann, spokesman for Greenpeace's gene technology campaign in Germany, dismisses the move as simply a way for the government to win time for consumers to become accustomed to gene technology. Greenpeace is fundamentally opposed to any deliberate release of GM organisms.

Beatrix Tappeser, coordinator of genetic engineering studies for the independent Öko Institute in Freiburg, an institute critical of agricultural biotechnology and which has the ear of the health ministry, says that "sufficient field trials of GM crops have already been conducted to demonstrate the dangers of gene transfer to the environment".

Without any details of the plan, says Tappeser, it is impossible to judge whether the programme suggested by Schröder could be helpful. But she points out that the environmental monitoring in such a programme should not be left to industry, or industry-financed researchers. "To win public confidence, monitoring needs to be conducted by all stakeholders," including environmental and consumer groups, she says. ■

'Dolly' team wins further patents

Peter Aldhous, London

The creators of Dolly the sheep have been awarded two more British patents on the 'nuclear transfer' technology that is the key to cloning. The patents should strengthen the position of the Dolly team, based at the Roslin Institute near Edinburgh, and of the companies that licensed its technology, in a highly competitive area.

In nuclear transfer cloning, a donor cell is fused with a recipient egg cell that has been stripped of its chromosomes. In January this year, the Roslin Institute won two British patents on cloning using donor cells that had been starved into a state of 'quiescence' — originally thought to be important for successful cloning.

By that time, however, rivals at the University of Massachusetts in Amherst had already been awarded a US patent covering cloning from non-quiescent cells, and had licensed this technology to Advanced Cell Technology, a company in nearby Worcester (see *Nature* 405, 610–612; 2000).

The new Roslin patents focus on

techniques to allow the donor nucleus to remain in contact with the recipient egg's cytoplasm for several hours before 'activating' the egg — starting its embryonic development using a pulse of electric current. "This is the other half of the nuclear transfer story," says David Earp, vice-president for intellectual property at Geron of Menlo Park, California, which has acquired the rights to exploit Roslin's cloning technology.

As the new patents do not depend on the donor cell being quiescent, they extend the Roslin team's claims. But just how powerful they prove to be remains to be seen.

The Roslin team argues that delayed activation can increase the efficiency of cloning, and in some species may be essential to achieve live births. But other groups say they can clone successfully without infringing the new patents. Goats, for instance, have been cloned by activating the egg during fusion with the donor cell (*Nature Biotech.* 17, 456–461; 1999). ■

Martian gullies tempt NASA to look for water

Michael Milstein, Wisconsin

Two strategies of the US space agency NASA may at last have paid off. A 'better, faster, cheaper' mission that also uses the 'follow the water' approach to Mars exploration has turned up what look like gullies freshly sculpted by water on frigid martian slopes where no liquid water should be.

At a news conference called hastily in Washington last week after word leaked out of a paper in the 30 June issue of *Science*, a team of planetary geologists studying images from the Mars Global Surveyor displayed sharp aerial photographs of red-rock washes and winding channels on the planet (see right). These features are strikingly reminiscent of deserts in the American southwest.

In some cases, the gullies cut through or blanket modern martian features such as sand dunes, suggesting that they formed very recently. If the photos shown had been of Earth instead of Mars, the scientists said, there would have been no doubt that water had carved the features.

"There's still a finite chance that they were formed some other way," said Michael Malin, head of Malin Space Science Systems, Inc., which built and operates the eagle-eyed camera aboard the Mars Global Surveyor. "But there's a high probability that they were formed by water."

In one fell swoop, the revelation has established compelling new scientific targets and generated renewed enthusiasm for what had become a rudderless NASA Mars programme beset by the loss late last year of two major missions, the Mars Climate Orbiter and the Mars Polar Lander.

At the same time, it appears to vindicate NASA's 'follow the water' strategy for Mars exploration, in which the agency has aimed to investigate the sinuous canyons and possible shorelines that may offer evidence of water in the planet's past, and which may be the best place to look for signs of martian life.

Ed Weiler, NASA associate administrator for space science, was quick to point out that the Mars Global Surveyor had been launched in 1996 under the same 'better, faster, cheaper' mandate that was now being blamed for fatal oversights in the navigation of the Mars Climate Orbiter and the design of the Mars Polar Lander.

"I'm just hoping the news today will pass the message that our Mars programme is not dead, in spite of all the rumours of its death a few months ago," Weiler said last week. "It is very, very pleasing to be talking about something positive for a change."

NASA must decide by July whether to send an orbiter or a rover to Mars in 2003. Malin, whose stock is sure to rise with these latest revelations, favours an orbiter with an even higher-resolution camera than on the



Water mark? Images from the Mars Global Surveyor appear to show gullies cutting through sand dunes.

Mars Global Surveyor; others now argue for a rover to follow the water on the ground.

Following the water on Mars may prove much easier than researchers had imagined if water is indeed emerging on the planet's cold, arid surface. Malin and co-worker Kenneth Edgett found about 120 places where water appeared to have gushed from the walls of impact craters, pits and large valleys.

Almost all of these are in the polar regions, usually on slopes shaded from the Sun. This places them in the coldest martian regions of all, where temperatures drop to -100°C , and where liquid water might seem least likely to appear.

But the two researchers say the cold may actually help by freezing dams of ice in place at the surface. Briny water building up behind the dams might finally burst out in torrents that then scour gullies in the slopes below before quickly vaporizing.

Many gullies end in what the researchers

called 'aprons' of sediment. These sometimes cover up modern sand dunes and lack impact craters, signs that they are "incredibly new" and may still be forming today, says Edgett.

But not everyone is convinced. Michael Carr of the US Geological Survey, who has spent decades looking for signs of water on Mars but is not part of Malin's team, says that water appears the most ready explanation. He adds that bursts of gas might also lubricate debris flows, and cautions against leaping to earthly conclusions on a different planet, especially one where water would not be expected to persist.

"Water is so difficult to understand, given the physical conditions on the surface of Mars, and we know those physical conditions very, very well — it's just simply too cold," says Carr. "I'm just having trouble reconciling the observation with the interpretation."

▶ http://www.msss.com/mars_images/

Astronomers fume over night light

Bill Triplett, Washington

This autumn, the main thing astronomers in part of the Washington area will see through their telescopes is irony. Discovery Communications, the parent company of the cable television channel specializing in science-orientated programmes and which sells telescopes in its shops, is planning to build a new global headquarters — with an illuminating twist.

The intended site is at Silver Spring, Maryland, not far from the Washington headquarters of the US space agency NASA. And the most prominent feature of the new building will be a 300-foot tower shooting a powerful beam of light into the night sky.

That is intolerable, say the area's astronomers, who are increasingly sensitive to the issue of 'light pollution', and are bemused that Discovery does not seem to be. "Isn't part of their charter to promote science and protect the environment?" asks Bob Gent, public affairs officer for the International Dark-sky Association (IDA). "Yet here they are destroying the night sky and ruining astronomy."

The night sky above the US capital is steadily deteriorating, as it is over most major cities around the world. "Light pollution in Washington is becoming more of a topic of concern for everyone," says John Settle, president of the Greenbelt



► Astronomy Club, an amateur group in Greenbelt, Maryland, which is near Silver Spring. "This is probably the second most polluted area [in the United States] after New York City, and this light will only degrade the quality of life for everyone around here. Its only purpose seems to be an ostentatious display of corporate power," he says.

Settle says that in the Greenbelt area, which is also home to an astronomy club composed of employees at NASA's Goddard Space Flight Center, star-gazers can see only second-magnitude stars. Spotting fifth- or sixth-magnitude stars, he says, needs more than an hour's drive.

There is no major astronomical research left in the Washington area, and the US Naval Observatory moved its heavy equipment and operations to Arizona more than 40 years ago.

Geff Chester, the observatory's spokesman, questions the reason for the planned tower light: "Why would anyone want to waste photons for something that's just going to light up the bellies of airplanes?" However, he doubts whether it will have much impact on the observatory's small-scale work.

But Gent says he has heard numerous complaints, not only from amateur astronomy clubs. "The American Astronomical Society has contacted me. So has the University of Maryland astronomy department," he says.

David Leavy, spokesman for Discovery Communications, says he is unaware of any complaints except for one letter from the IDA. But he admits that a recent article in a local paper about the plans for the building has made the company sensitive to the concerns.

So far, he says, no changes have been made to the design. "But as we go forward we'll be consulting with the organizations to make sure this structure is something... that will be environmentally sound and will not cause any celestial inconvenience." Rest assured, says Gent, the IDA will be watching. ■

Educated US public get more wary of genetic engineering

Paul Smaglik, Washington

Well-educated Americans have become less keen on genetic engineering over the past five years, according to a survey carried out by the National Science Board (NSB). This is despite a general increase in acceptance of the topic in the previous decade.

The proportion of US adults with bachelor degrees who responded that the potentially harmful results of genetic engineering either 'slightly' or 'significantly' outweighed the benefits is rising, from 20% in 1995 and 24% in 1997 to 29% in 1999.

The results, published last week in the NSB's *Science and Education Indicators 2000*, point to a growing suspicion of the biotech industry by some sectors of the public, at a time when the industry, prompted by the full sequencing of the human genome (see page 983), is set for greater financial prosperity.

Negative perception of genetic engineering across all US adults has remained slightly less volatile. Adults with less than a college education are still more suspicious of genetic engineering as a whole, while adults who have completed high school or college or who follow progress in medical research have become more suspicious of it since 1995.

But some are sceptical. Dan Eramian, spokesperson for the Biotechnology Industry Organization (BIO), says it "runs counter to every national survey I've seen".

Eramian says the public has little idea of what 'genetic engineering' means. The term encompasses cloning, gene therapy and genetically modified (GM) foods, among other transgenic technologies. Those polled may have responded differently if they had been asked about each specific technology, according to the report.

But Jeremy Rifkin, president of the Foundation on Economic Trends in Washington, a group that opposes GM foods, says that the more BIO promotes the applications of genetics to promote its image, the more people appear to become opposed to it.

Rifkin thinks that the falling support of genetic engineering by educated and informed people reflects uncertainty, rather than panic. "People are saying we don't have enough information to show that this is safe," he says.

Dorothy Nelkin, a professor of sociology at New York University who specializes in science and law, says she is not surprised at the results. People in grass-roots organizations who oppose GM food, she says, tend to be both highly educated and informed of progress in science and medicine.

Nelkin suspects that the public fears industry taking over science more than it fears the science itself. "Commercialization enhances mistrust," Nelkin says. ■

♦ <http://www.nsf.gov/sbe/srs/seind00/start.htm>

Respondents who argue that the harmful results of genetic engineering 'slightly' or 'strongly' outweigh benefits

	1985	1990	1995	1997	1999
All adults	39	37	35	36	38
Male	35	34	32	21	33
Female	42	41	38	40	42
Less than high-school graduate	36	32	42	44	36
Baccalaureate and higher	31	28	20	27	26
'Attentive' to medical research	36	30	27	30	36

Call for more accelerator research

Quirin Schiermeier, Munich

Luciano Maiani, director-general of the Geneva-based European Laboratory for Particle Physics (CERN), last week called on the organization's 20 member states to put in place a "vigorous programme of accelerator research and development" to help maintain CERN's leading position in particle physics in the long term.

CERN's current core activity, the search for the still elusive Higgs particles, will continue, Maiani told the CERN Council last week. Higgs particles, if they exist, are

predicted to be in the range of CERN's new Large Hadron Collider (LHC), experimental work on which is to start next year.

But funding for future work, including the Compact Linear Collider (CLIC) and research into high-intensity proton accelerators aimed at producing a second-generation neutrino beam, is "barely sufficient," he said. These projects currently receive around 1% of CERN's resources. This was acceptable during the construction phase of the LHC, Maiani said, but now "it is time to start a gradual increase". ■

Patent suit on Alzheimer's mouse rejected...

Rex Dalton, San Diego

Neuroscientists worldwide can continue to enjoy access to an important transgenic mouse used for research into Alzheimer's disease, following the rejection of a patent-infringement claim against the institution that distributes the mice.

Elan Pharmaceuticals, Inc., based in Ireland, filed a suit against the Minnesota-based, non-profit Mayo Foundation for Medical Education and Research last year, alleging that two of its US patents for transgenic mice were being infringed by the foundation's strain of transgenic mice (see *Nature* **404**, 319–320; 2000). If successful, the move could have blocked the Mayo Foundation's distribution of the mice.

But on 15 June, Judge William Alsup of the US District Court in San Francisco dismissed Elan's lawsuit. He ruled that the company's two patents were invalid, as they involved technology already included in an earlier patent that the Mayo Foundation was licensed to use.

"This is really fantastic news," says Steve Younkin, a neuroscientist at Mayo's facility in Jacksonville, Florida, where its Alzheimer's research is concentrated.

Elan intends to appeal against the decision. "We are confident the patents are valid and the court will reverse the decision," says Libby Murphy, Elan's executive vice-president for intellectual property and legal affairs. An Elan spokesman adds, "It is unfortunate we have to defend our intellectual property against the Mayo Foundation, as we have worked closely with them in the past and hope to do so in the future."

The Mayo Foundation's transgenic mice were developed and distributed using technologies licensed from the University of Minnesota and a small Kansas company, the Alzheimer's Institute of America.

Judge Alsup ruled that the Alzheimer's Institute's technology, patented by neuroscientist Michael Mullan of the University of South Florida (see right), preceded Elan's patent disclosures. "The Mullan patent application disclosed the same recipe for making transgenic mice as was later disclosed in the [Elan] patents," Alsup wrote in his eight-page ruling.

Karen Boyd, the Mayo Foundation's attorney with the law firm of Fish & Richardson, says the court ruling "gives researchers at the Mayo Foundation, other academic institutions and biotechnology companies the opportunity to continue research".

The transgenic mice of both organizations are different, but they both involve engineering the mice to have a human genetic mutation that is linked to a build-up of amyloid protein leading to neurodegeneration in the brain.



Threat lifted: Alzheimer's patients stand to benefit from research with the Mayo Foundation's mice.

The Mayo Foundation has been distributing its mice to academic institutions at nominal costs, although it has been charging some pharmaceutical companies as much as \$850,000 for a breeding group.

Elan is known for exerting tight control over the distribution of its products to academic researchers and other drug companies.

Two disputed patents were acquired when Elan purchased Athena Neurosciences, Inc. in 1998. If the court ruling stands after Elan's appeal, the two patents would be permanently invalidated, and the US Patent and Trademark Office would withdraw them. A decision by an appeals court in Washington usually takes a year. ■

...but controversy over rights lingers on

A patent at the centre of a heated court battle over tools used for research into Alzheimer's disease (see above) itself emerged from a controversial arrangement between an expatriate British scientist and a public university in Florida. In addition, a Swedish research team argues that it should have shared in the rights.

The patent covers the genetic sequence of a human mutation that causes Alzheimer's disease and its use in transgenic mice (*Nature Genet.* **1**, 345–347; 1992). It was invented by Michael Mullan, a neuroscientist who trained at Imperial College in London.

Mullan says he received the rights to the discovery under a deal with the University of South Florida in Tampa.

Mullan left England in 1992 and moved to the university having, he says, become disheartened with how he was treated when Imperial College sought to capitalize on his scientific discoveries.

Mullan says his terms of employment at South Florida allowed him to own the patents on any scientific discoveries he made

within about six months of his arrival in Florida.

During this time Mullan and his colleagues sequenced the gene in question. Mullan sold the patent rights to the Alzheimer's Institute of America, a company set up by Kansas venture capitalist Ron Sexton, who had been funding Mullan's work. The Mayo Foundation later licensed the technology from Sexton's firm.

Kenneth Preston, the University of South Florida's director of patents and licensing, acknowledges that initially "there was an agreement not to apply rules and procedures regarding intellectual property" to Mullan. He also accepts that, as a result, a discovery that would normally have been owned by the university ended up in private hands.

Mullan, who now directs the university's Roskamp Institute, says that he did everything "above board" and "in full view of everyone". He says that the university's officials "felt sympathetic" to him for having been badly treated by Imperial College. The college declines to comment on the issue.

In 1991, Imperial College sold the rights to the first Alzheimer's mutation that Mullan helped to discover (see *Nature* **353**, 844–846; 1991) to Athena Neurosciences, Inc., which was later bought by Elan, the company at the heart of last week's court hearing.

Another twist is the fact that the DNA used by Mullan to sequence the mutation at South Florida was provided by a Swedish research team headed by Bengt Winblad, chief of clinical neuroscience at the Karolinska Institute in Stockholm.

In publishing the findings, Mullan shared co-authorship with the Swedes in the 1992 *Nature Genetics* article. But he listed himself as sole inventor on the US patent, which was granted in 1995.

Lars Lannfelt, a neurogenetics professor on Winblad's team who unsuccessfully fought for recognition on the patent, claims that he and his colleagues should not have been omitted. Mullan denies this, arguing that the Swedish researchers "didn't have a position" as inventors, because they only provided DNA, and made no intellectual contribution. **R. D.**

UK studies ways of paying to join Chile observatory

Natasha Loder, London

British astronomers are bracing themselves for the project cuts that may be needed if Britain is to join the European Southern Observatory (ESO). This is the eight-member organization that operates the recently opened Very Large Telescope on Mount Paranal in Chile.

The Particle Physics and Astronomy Research Council (PPARC) has already opened negotiations with ESO officials in Munich over possible membership (see *Nature* 405, 382–383; 2000). It believes that, like space and particle physics before it, the price of world-beating astronomy is now beyond the reach of individual countries.

But the cost of joining the ESO will be high, with an entry fee of at least £60 million (US\$90 million) and a further £12 million every year. PPARC officials say that entry depends both on whether the government will provide extra funding in the next spending round, and whether it can find about £7 million a year from existing operating costs.

There was little dispute over the principle of joining the ESO among 150 astronomers who came together last week at a meeting in London organized by the Royal Astronomical Society. If Britain does not join the ESO, one warned, it would create a “lost generation” of young astronomers.

The question now is how to pay for it. According to Ian Corbett of PPARC, cost-cutting options include withdrawing from the Anglo-Australian Telescope in New South Wales, Australia, and cutbacks at the Isaac Newton Group of Telescopes in Spain and the United Kingdom Infra-Red Telescope and the James Clerk Maxwell Telescope in Hawaii.

One possibility is that Britain might bargain for a reduction in the annual fee with a payment in kind of European access to UK facilities. Several possibilities were raised, including the Visible and Infrared Survey Telescope for Astronomy (VISTA) in Chile.

But Jim Emerson, head of the consortium of UK universities that recently won the bid for the VISTA telescope, pointed out that VISTA is not owned by PPARC but by the consortium, so it “can’t just be given away”.

Catherine Cesarsky, director general of the ESO, says discussions about trading are premature and that, if any trades are made, they would be only a small part of the overall deal. ■

Pasteur Institute to abandon departmental structure?

Declan Butler, Paris

The new head of the Pasteur Institute, one of France’s leading biomedical research institutions, has begun a major shake-up.

Philippe Kourilsky took over as director-general of the private institute in Paris earlier this year (see *Nature* 401, 630; 1999). In a meeting last week with its 1,100 scientists, he outlined a series of plans intended to reinforce central management and remove dead wood. In particular, he is disbanding the institute’s vertical system of departments in favour of a more horizontal system organized around projects.

Over the past few months Kourilsky has put in place a new central management structure and recruited heavily. Eleven new divisions have been created corresponding to key activities, including scientific careers, multidisciplinary programmes, technology platforms and relations with France’s public research organizations.

The technology platforms division, headed by tuberculosis geneticist Stewart Cole, is intended to provide scientists with centrally run services in genomics, proteomics and bioinformatics. One-time venture capitalist Christian Policard, formerly a member of the board of pharmaceutical company Sanofi-Synthelabo, will head the division of industrial affairs.

The institute needed to be “professionalized”, says Kourilsky, adding that he was influenced by a three-year stint in the private sector as director of research at Pasteur Mérieux Connaught.

The most drastic step, announced in January, is a rule stating that laboratories more than 12 years old must make a case for their continued existence or be closed. The measure is being applied retroactively and will affect almost half of the institute’s laboratories.

Affected laboratories have until the end of October to put forward proposals. Kourilsky warns that laboratories that propose to continue as they are “run the risk of having their applications rejected”. Staff who lose their laboratories will be redeployed.

An international call for proposals will be made in September to create new laboratories of five to six researchers led by young scientists. Ten such laboratories, which will have a lifespan of five years, will be created next year, and the ambition is to create 50 within five years. At present, the institute has around 100 laboratories, each with an average staff of 20 to 25.

Kourilsky adds that the current systems for evaluating individual laboratories and staff has suffered from “inbreeding”. In



All change: Kourilsky (inset) seeks to increase flexibility at the Pasteur Institute.

future evaluation will be largely carried out by external panels.

A ‘foresight’ exercise, to be completed by the end of the year, is likely to pave the way for abolition of the institute’s ten departments, with projects and programmes between laboratories becoming the main vehicle for implementing research strategy.

Harold Varmus, former director of the US National Institutes of Health and now president of the Memorial Sloan-Kettering Cancer Center in New York, will chair a new eight-member External Committee of Strategic Orientation, intended to provide outside assessment of the institute’s general directions.

Kourilsky sees the flexibility allowed by this move as critical to his goal of refocusing the institute on its core mission of combating infectious diseases. At the same time, he has introduced a programme to fund projects spanning the institute’s laboratories, which he says have become fragmented.

To force this change, he froze 30% of laboratory funding at the beginning of the year, to fund a new series of projects that must include at least three of the institute’s laboratories. Three-quarters of the laboratories have responded to the call for proposals, and the winners will be announced next month.

So far, researchers at Pasteur seem to be giving Kourilsky the benefit of the doubt; as the institute is private, it has greater freedom to impose reforms. “Change was needed,” says one scientist. But there is some scepticism as to how the structural changes will work in practice, given their volume and the proposed speedy timetable. ■

Los Alamos tapes 'never left lab' Senate told

Washington US energy secretary Bill Richardson told a Senate hearing last week that the Federal Bureau of Investigation had found no evidence to suggest that computer tapes containing nuclear secrets — which went missing from the Los Alamos National Laboratory, and later turned up next to a photocopier — had ever left the laboratory (see *Nature* 405, 877; 2000).

"So far there is no evidence of espionage, nor is there evidence that the drives have ever left the Los Alamos X Division," Richardson told the Senate Armed Services Committee last week. The disk drives are thought to have disappeared at the end of March, but their absence from a secure vault was not detected until 7 May.

The chair of the committee, Senator John Warner (Republican, Virginia), said he planned to introduce legislation to look at whether responsibility for the nuclear weapons programmes — including the weapons laboratory — should be given to the defence department or to an independent agency.

Russian unions protest over axed ministry

Moscow The Russian coordinating committee of the scientific trades union organizations has called for a national day of protest next week (4 July) over the removal of the science ministry in a recent cabinet reshuffle.

The committee has told president Vladimir Putin that scientists are shocked by the move, which means that "there is no one in the cabinet who is directly and solely responsible for Russian science".

The first demonstrations will be held in Moscow near the Kremlin and the State Duma, the lower chamber of the Russian parliament. Members from across the country have offered to mount local protests in support of the committee.

Swedish centre to focus on biosphere

Stockholm Lund University in Sweden is to set up a Centre for Biosphere-Geosphere Studies. The centre is intended to act as "a powerful base for modern basic research and education with clearly defined applications towards important global and regional environmental problems".

The centre's main research objective will be to promote an improved understanding of the processes linking the geosphere and biosphere. It will also seek to ensure collaboration and coordination between national and international research projects

in biosphere-geosphere interactions. The centre will be associated with a new Geocentrum being constructed in Lund, and should be ready for use in less than two years.

Budget funding boost for New Zealand science

Sydney Science in New Zealand has received its largest funding increase in a decade, as part of the Labour-led coalition government's first budget (see *Nature* 401, 106; 1999). In a shift from the market focus of its predecessor, the government has raised the research and development budget by almost 10%, to a total of NZ\$474 million (US\$222 million).

Most of the extra funding will be targeted to support research in industry, although business has attacked the government's scrapping of an election promise to introduce tax incentives in favour of grants to small companies. Already under strong financial pressure, universities have been told to direct their new funds to stabilizing the growth in tuition fees, which was a major issue in last year's election.

UK on target for Kyoto carbon dioxide cuts

London A report on five European countries shows that only the United Kingdom is on track to achieve its Kyoto Protocol target for reduction in carbon emissions. The report, released by the Pew Centre on Global Climate Change, also looked at Germany, the Netherlands, Austria and Spain. It says that all need to put new measures into place if they are to reach their reduction targets.

The report adds that Germany is close, but will have difficulty reaching its target of 21% below 1990 levels. The Netherlands, meanwhile, has a target reduction of 6%, but current CO₂ emissions have increased by 17%. The authors of the report include the former UK environment minister John Gummer, now chairman of an environmental consulting firm.

Stone marks site of Heisenberg's inspiration

Munich The remote North Sea island of Helgoland became the scene of science history 75 years ago, when Werner Heisenberg made the major breakthrough in his theory of quantum mechanics there. This month, a memorial stone was erected on the island to celebrate the anniversary of the German physicist's breakthrough in June 1925.

He had been staying on Helgoland, one of Germany's most northerly points, to escape the hay fever that affected him on the mainland. Heisenberg, who died in 1976, was awarded the Nobel Prize for physics in 1932 for his work on quantum mechanics and for



Heisenberg monument: Gerd Buschhorn (left), director of the Max Planck Institute for Physics in Munich, and Frank Botter, the Bürgermeister of Helgoland were among those at the unveiling.

elaborating the principle of indeterminacy in nuclear physics.

Research council fined for technician's death

London Britain's Medical Research Council (MRC) has been fined £25,000 (US\$37,000) over the death last year of a lab technician in its Human Genetics Unit in Edinburgh. Jimmy Graham, aged 51, was asphyxiated when decanting liquid nitrogen from a storage vessel. The prosecution at Edinburgh Sheriff's Court followed a report to the procurator fiscal by the Health and Safety Executive (HSE).

The HSE has welcomed the fine as recognition of the severity of the event. The MRC, which pleaded guilty to breaching health and safety regulations, has expressed regret over the death, stating that "the safety of our staff is our first concern". It adds that new procedures have been put into place to ensure that such an event does not occur again.

Gates to Japan open to Canadian biomedics

Montreal Canadian biomedical postdocs will be eligible to train in Japanese national institutes under one of the world's most attractive fellowship programmes, thanks to an invitation from the Japanese Science and Technology Agency to the Canadian Institutes for Health Research (CIHR).

The CIHR was launched on 7 June as Canada's main federal funding agency for health research (see *Nature* 405, 722; 2000). Until now, Canadian applicants to the programme had been handled by the Natural Sciences and Engineering Research Council. Canada hopes that enough applicants will pass through the screening system to reach the current quota of 14 long-term and four short-term fellows.

Feathers fly in Beijing

The question of whether birds evolved from dinosaurs arouses strong opinions. Rex Dalton reports on a scientific meeting that at times bore more resemblance to a political sparring match.

When ornithologists and palaeontologists rolled into Beijing earlier this month for the fifth quadrennial meeting of the Society of Avian Paleontology and Evolution, it seemed like a perfect opportunity to try and resolve the long-running debate over whether birds evolved from dinosaurs.

Northeast China has produced a wealth of bird and dinosaur fossils in recent years, and these have been used to explore the links between the two groups. At the meeting, Chinese scientists showed off some newly discovered specimens, which might help to answer important evolutionary questions.

But by the close of the meeting, hosted by Beijing's Institute of Vertebrate Paleontology and Paleoanthropology (IVPP), the divisions between those who believe birds evolved from dinosaurs and those who disagree appeared greater than ever. Several attendees were disappointed that progressive scientific debate was stifled by entrenched opinions, raised voices and strident words. "I think a lot more interesting issues could have come up, instead of going over the same old tired stories, like the origin of flight," says John Hutchinson, a doctoral student at the University of California at Berkeley. "It is like the field hasn't moved on."

"The time is now to stop debating and proceed with learning about the biology of the origin of birds," agrees Richard Prum, an evolutionary biologist at the University of Kansas in Lawrence. But when Prum argued at the meeting that birds' feathers have the same



Bone China: this skull of a feathered dinosaur, *Caudipteryx zoui*, is one of the exciting new fossils emerging from northeast China.

evolutionary origin as the hair-like integumentary filaments seen on many dinosaur fossils, he was accused by Storrs Olson, head of ornithology at the National Museum of Natural History in Washington DC, of engaging in "ideological mumbo jumbo".

Olson is a leader in the camp that believes that birds evolved separately from dinosaurs. He rejects the phylogenetic analysis used by Prum and others to build evolutionary links between birds and dinosaurs on the basis of shared characteristics. Throughout the meeting, Olson and like-minded scientists wore badges saying "Birds are Not Dinosaurs".

But this 'BAND' group was not alone in adopting tactics that had more in common with political point-scoring than scientific discourse. When Alan Feduccia, a palaeobiologist at the University of North Carolina at Chapel Hill, gave a talk on why the theory of bird evolution from dinosaurs should be rejected, his argument was likened by some opponents to those of creationists. Chris McGowan, curator of palaeobiology at the Royal Ontario Museum in Toronto, said cuttingly that he had not enjoyed such a performance since he last heard a talk by Duane Gish, senior vice-president of the Institute for Creation Research in Santee, near San Diego.

But between such acrimonious exchanges, new lines of research did emerge. Mary Schweitzer of Montana State University in Bozeman described how she has looked for evidence of feathers on a Mongolian dinosaur fossil, *Shuvuuia deserti*, using fluorescently labelled antibodies for the protein keratin. Schweitzer detected β -keratin, the basic component of most feathers, but significantly she found no evidence of α -keratin, which does not occur in feathers.

Hutchinson, who studies at Berkeley with Kevin Padian, a leading proponent of the idea that birds evolved from dinosaurs, presented data on hind-limb evolution. Hutchinson compares bones of fossils and existing species for scarring to determine where muscles attach. He uses this information to generate a computer model of the biomechanics of hind limbs. As an example, he presented preliminary data on how much hind-limb muscle mass would be needed to support a *Tyrannosaurus rex*. Hutchinson hopes that the modelling process will be useful for exploring the links between birds and dinosaurs.

Kevin Middleton, a doctoral student in evolutionary biology at Brown University in Providence, Rhode Island, described his studies of the feet of birds and dinosaurs. He has focused on the evolution of the first digit, or hallux, the orientation of which is used to help classify species as birds or dinosaurs. The bone is reversed in perching birds, but Middleton has produced evidence that it is not only reversed, but also rotated along its axis. "I'm fairly confident the hallux is much more complex than we thought," says Middleton.

"These are examples of the next generation of research that will provide tremendous progress," says Prum. And the specimens being uncovered in northeast China are likely to be key to this progress. During the meeting, one IVPP scientist, Fucheng Zhang, presented tantalizing slides of about half a dozen undescribed species.

But if the Beijing meeting is anything to go by, any resolution of the debate over the origin of birds may take years to emerge. When the proponents get together in the same room, they seem to generate more heat than light. ■

Rex Dalton is *Nature's* West Coast US Correspondent.



Olson: BAND leader rejects the bird-dino link.

REX DALTON

Africa needs locally trained soil specialists to improve land use

Sir—I commend and support Willy Verheye's suggestion in Correspondence¹ that improved road systems to markets and fair price-setting will allow inland African farmers to compete better with those who sell cheaper imported foods. However, his corollary suggestion of using the "large untapped land-use potential" in this part of the world to increase food production is contrary to sensible notions of development and land management, for several reasons.

First, the few data available on land resources are unreliable. The FAO World Resources Report² cited by Verheye was intended to highlight the global-resources aspects of the 28 major soil association groups and of their continental distribution on a 1:25 million map. Only one page (page 28) and one table (Table 11, containing some obvious errors) are devoted to land-resources reserves, and no details are given for how the data for the developing countries were obtained. The accuracy of the data at the scales used for the preparation is doubtful: each map unit includes various soil types and possible land uses². Further, many specialists would dispute the data on cultivable reserves—for example, there is one estimate³ of possible arable expansion of 500 million ha for the whole world, which differs greatly from the 622 million ha for Africa and 696 million ha for South America in the FAO report².

Second, current attitudes about nature conservation and development are strongly against further unchecked deforestation, which is likely to speed up soil erosion and land degradation, as well as having negative effects on biodiversity and possibly on global climate change. The remaining forests in Africa and elsewhere must be protected from further destruction, in order to promote a global sustainable geo-ecosystem. The World Bank is now rightly promoting economic development with complementary environmental conservation. African governments should follow this lead.

Yet Verheye is correct to say that solving Africa's food-production problem is, in principle, relatively simple. Food security for the growing population does not necessarily mean expansion of cultivated soil areas, but rather improved soil care and nutrient management on existing arable soils. Despite the increased production in recent decades, the average yields of cereals in Africa are still only one third of those in Europe and North America (ref. 3 and the FAO annual yearbooks). Yields per hectare of most crops in developing countries are only a fraction of what they could be under

proper soil care and nutrient management, as shown by several Israeli and other demonstration projects in developing countries. Good soil care and land-use management involves soil testing, rotation, terracing and appropriate tillage, integrated with nutrient and moisture management⁴. Soils in practically all African countries are depleted of nitrogen, phosphorus and potassium. This nutrient depletion is probably increasing because of monocropping and insufficient resupply^{5,6}. The vicious cycle of marginal inputs producing marginal yields resulting in marginal living standards must be broken.

In addition to improving the transport system, as Verheye suggests, the current subsidy and overseas aid should be used to provide the fertilizers and improved planting materials that are needed. Many locally trained soil and extension specialists living in the region are needed to transform the economy from one of small rural farmers to one of market food production^{7,8}. At least ten dedicated extension soil scientists per million of rural population with problem-solving skills would be a good start.

Dan H. Yaalon

Institute of Earth Sciences, Hebrew University, Jerusalem 91904, Israel

1. Verheye, W. *Nature* **404**, 431 (2000).
2. FAO World Soil Resources Report 66, *An Explanatory Note* (FAO, Rome, 1991).
3. Biswas, M. R. *AMBIO* **23**, 192–197 (1994).
4. Yaalon, D. H. & Arnold, R. W. *Soil Science* **165**, 5–12 (2000).
5. Henao, J. & Baanante, C. *Estimating Rates of Nutrient Depletion of Agricultural Lands of Africa* (International Fertilizer Development Center, Muscle Shoals, 1999).
6. Hartemink, A. E. *Geoderma* **75**, 215–229 (1997).
7. Lal, R. *Soil Science* **165**, 57–72 (2000).
8. Yaalon, D. H. *Soil Science* **161**, 3–8 (2000).

Did Greeks beat Chinese on blood circulation ...

Sir—I was drawn to read P.-L. Chau's letter about the Chinese discovering blood circulation¹ by the provocative title "Chinese beat Harvey on blood flow" on the contents page. Chau quoted science historian Joseph Needham² as saying that, in the medical treatise *Su Wên* (part of *The Yellow Emperor's Classic of Internal Medicine*), "Chhi Po says that 'the flow in the tract and channel runs on and on, and never stops; a ceaseless movement in an annular circuit...'. Clearly the circulation of the blood and chhi was standard doctrine [in the second century BC]".

But who is Chhi Po? According to Paul Unschuld³, a medical historian and authority on Chinese medicine, Chhi Po or Qí Bó, the most important interlocutor of the Yellow Emperor in *The Yellow Emperor's Classic of Internal Medicine*, is none other than Hippocrates.

Qí Bó is a man who has no historical

background in Chinese history or mythology. This fact, together with the Hàn-period pronunciation of his name, allows speculation that the fame of the Greek physician reached China two centuries after his death—to the extent that he is quoted as a living authority in medical textbooks of the time.

Tsung O. Cheng

Department of Medicine, The George Washington University Medical Center, 2150 Pennsylvania Avenue, NW, Washington DC, USA

1. Chau, P. L. *Nature* **404**, 431 (2000).
2. Needham, J. & Lu, G. D. *Celestial Lancets* (Cambridge Univ. Press, Cambridge, 1980).
3. Unschuld, P. U. *Chinese Medicine* (Paradigm, Brookline, 1998).

... or was 'blood as the river of life' just poetic?

Sir—Once again, the idea that the Chinese discovered the circulation of the blood before Harvey has been put forward, this time in Correspondence¹. Joseph Needham² does indeed quote passages from Chinese medical classics that, according to his translation, suggest the Chinese knew of the circulation of 'vital energy' (*chhi*, or *ch'i*, or *qi*) and that the heart acted as a pump. The problem is that Needham's translation of the passages is not accepted by other experts.

Paul Unschuld, for example, translates the same passages in a way indicating that the Chinese knew neither of the circulation of the blood (or *qi*) nor that the heart acted as a pump^{3–5}.

Unfortunately, with a little stretch of the imagination and encouraged by the will to prove a point, one often can see 'proof' or suggestions of knowledge of the circulation when such proof or suggestions are not there. For example, the author of *Peri kardies*⁶ in the Hippocratic Corpus says that the aorta and the pulmonary arteries "are the springs of the nature of man, and the rivers there move throughout the body by which the body is watered". Dante⁷ says "And the blood, which fills the veins, runs toward the heart which calls it."

Should we therefore conclude that Dante and the author of *Peri kardies* had at least an inkling about circulation? Hardly.

Plinio Prioreschi

Department of Pharmacology, Section of History of Medicine, Creighton University, 2500 California Plaza, Omaha, Nebraska 68178, USA

1. Chau, P. L. *Nature* **404**, 431 (2000).
2. Needham, J. & Lu, G. D. *Celestial Lancets* (Cambridge Univ. Press, Cambridge, 1980).
3. Unschuld, P. U. *Nan-ching: The Classic of Difficult Issues* 14 (Univ. of California Press, Berkeley, 1986).
4. Unschuld, P. U. *Medicine in China: A History of Ideas* 371 (Univ. of California Press, Berkeley, 1985).
5. Prioreschi, P. A. *History of Medicine* Vol. 1, 2nd edn, 136 (Horatius Press, Omaha, 1996).
6. Hurlbutt, F. R. *Bull. Hist. Med.* **VII**, 1104–1113 (1939).
7. Alighieri, D. *Rime* **VI**, CIII, 45–47 (Longo, Ravenna, 1995).

A question of compliance

How an epidemic was overcome by criminalizing the patients.

From Chaos to Coercion: Detention and the Control of Tuberculosis

by Richard Coker

St Martin's Press: 2000. 262 pp. \$27.95

Thomas Dormandy

The tuberculosis epidemic that struck New York City in the late 1980s may have been no more than a blip in the conquest of the disease, and might even be represented as the triumph of a vigorous public-health response. After an alarming upsurge and an even more alarming increase in multi-drug-resistant cases, the epidemic was under control by 1996. Nevertheless, such a representation would be misleading. Tuberculosis has gone into a latent phase many times before, only to re-emerge when conditions were favourable; and major epidemics have often been heralded by blips not unlike the New York outbreak. Also, the suppression of that outbreak was achieved at a cost which, on a larger scale, would probably be unacceptable in modern Western society. Nothing suggests that the underlying causes of the epidemic have been removed, and much suggests the opposite.

From a historical viewpoint, it is impossible not to be struck by echoes from the past. From the start, the New York public-health response had two ostensible aims — first, to treat patients; and second, to protect the public. Just as happened with sanatoria 100 years ago, the verbiage focused on the first, the action on the second.

The disparity in the two aims was fuelled by the social context. Tuberculosis had been tacitly accepted as a regrettable but unavoidable feature of poverty, overcrowding, ignorance, homelessness, drug abuse and social alienation. It had no business venturing outside these areas. Once again, the organism that threatened to transgress social barriers showed characteristics not seen before. For 50 years tuberculosis had been a scourge, but a curable scourge. The New York outbreak established the link between the mycobacterium and the AIDS virus; and multi-drug-resistance made the disease potentially as lethal as it had been in the days of John Keats.

The verbal response, too, uncannily echoed the past. For, contrary to the wisdom of hindsight, tuberculosis had always been treatable. At different times there was sunlight, high altitude, sea-bathing, rest, exercise, calcium, blood-letting, the pumping of air into various body cavities, the wholesale resection of ribs, the crushing of nerves, the removal of lungs ... the list could continue. Now there was DOT (directly observed



Necessary evil: screening the population for tuberculosis forms part and parcel of its suppression.

therapy), the whiplash sound inspiring instant confidence. (DOT, as its full name suggests, involves face-to-face monitoring by a health-care worker that the patient is taking the treatment prescribed.) But DOT, or any kind of treatment, requires a measure of willing collaboration — or compliance, to adopt the new terminology — on the part of patients. This raises an age-old dilemma. A patient with heart disease or cancer who chooses not to comply usually endangers nobody but him- or herself. A tuberculous patient, likely, by the very nature of the disease, to be a feckless, witless, jobless, homeless individual, can endanger thousands of entirely respectable fellow citizens. This, in the world's most health-conscious (as well as freedom-loving) country, was intolerable.

Hardly anybody protested at new legislation that enabled health authorities to coerce non-compliers. George Comstock, an eminent epidemiologist, expressed the widely held view that "people who deliberately or carelessly infect other people should be treated as criminals ... They're just as dangerous as the guy who goes about shooting off his pistol randomly." A meeting sponsored by the New York Academy of Medicine concluded that "recalcitrant patients should be detained in single-disease institutions until they were cured [or, presumably, died]". As happened time and again in the past, the transition from hating tuberculosis to hating the tuberculous was astonishingly swift.

In one other and more heartening respect, the New York epidemic resembled bygone crises. At the grandiose tuberculosis conferences of the late nineteenth and early twentieth centuries, pressure and persuasion were never entirely successful: one or two individuals usually braved the opprobrium

of their colleagues and openly expressed dissent. *From Chaos to Coercion* by Richard Coker, a London physician, is in this splendid tradition.

The author witnessed the New York epidemic, and spoke to many of the protagonists. Like his predecessors, he cannot provide answers. But, almost equally important, he asks the right questions. One such question is, not whether the coercion of non-compliers was successful, but whether it was necessary. (If it was necessary, was it necessary in order to raise the credibility of torpid public-health authorities or to stamp out the disease?) Does DOT really work? (The truth does not always speak in acronyms.) Why is a disease of poverty and overcrowding increasing in the richest city in the world? (It still is.) Why do large numbers of individuals with tuberculosis fail to comply with treatment? (This is a far more penetrating question than what should be done with the non-compliers.)

Coker's book is lucid and intelligent, but not always an easy read. It is part reportage, part socio-political tract, part analysis of national characteristics. Tuberculosis has been called the perfect expression of an imperfect society, and no serious writer can ignore its social aspects; but this makes the need for organizing the material all the more important. However, this is no more than a cosmetic blemish on what is a fascinating, topical and challenging work.

Still in a historical context, tuberculosis has spawned thousands of more or less learned medical tomes, now both unreadable and unread. It has also inspired a few literary masterpieces by sufferers or their relatives. One of the patients Coker mentions, Paul Mayho, a victim of multi-

drug-resistant tuberculosis, has written such a first-hand account, *The Tuberculosis Survival Handbook* (XLR8 Graphics, London, 1999). It is a slender but revelatory volume whose jokey title does it no justice; it should be read by anyone involved with this still terrible disease. ■

Thomas Dormandy is at 16 St Albans Road, London NW5 1RD, UK.

Reactions of a chemical kindred

Candid Science: Conversations with Famous Chemists

edited by István Hargittai
Imperial College Press: 2000. 516 pp.
£25 (pbk)

Gautam R. Desiraju

Science, one would like to think, is always candid, and here we have a collection of candid scientists who speak at length about themselves and their work. Just about every one of the 40-odd chemists interviewed by István Hargittai could be termed famous, at least within the chemical community. Around half are Nobel laureates, and the editor has chosen well in that, taken together, the conversations provide a broad overview of the development of chemical thought in the second half of the last century.

The interviews have appeared individually in *The Chemical Intelligencer*, a journal which itself owes much to Hargittai as its founding chief editor. The present book is a lightly edited collection of this material. The interviews are not published in any particular sequence, but, having read them individually in *The Chemical Intelligencer*, I must admit to a feeling of satisfaction in seeing them all together. The whole is greater than the sum of its parts.

A third or so of those interviewed are native-born Americans, another third are scientists of European extraction, mainly Jewish, who migrated to the United States around the tumultuous times of the Second World War, while the remaining interviewees spent a large part of their working lives in the countries of their birth, whether in Europe or elsewhere.

The American influence is pervasive in this book, as it is in modern chemistry. Although the editor does not mention the circumstances that led to his choice of these particular chemists, the very complete picture that emerges only highlights the dominance of the US academic-industrial synergism in establishing trends and setting priorities in contemporary chemical research. Carl Djerassi comments about the fashion orientation of American chemistry, but is quick to admit that he speaks as an



A cornucopia of chemists: words from Pauling through to Zewail.

outsider. This brings one to the next question — does the outsider have in-built advantages as a researcher? According to Erwin Chargaff, each pioneer is *eo ipso* an outsider. Going by the interviews here, one can safely turn the aphorism around.

Indeed, a significant theme of this book is what it takes to make an outsider into a pioneer. Gertrude Elion, Paul Scheuer, Vladimir Prelog, Michael Dewar, Roald Hoffmann, Herbert Brown, George Olah, Eiji Osawa and Ahmed Zewail have much to say on this matter, in addition to Chargaff and Djerassi. These chemists come across clearly as outsiders with respect either to the establishment, to their adopted country, to society and its conventions, or more poignantly, with respect to their families. But in every case, self-perception as an outsider seems to have triggered vital chemical reflexes. Can internal unrest spark scientific imagination? Clearly, yes, although other equally stimulating reasons are apparent from the conversations with, say, John Pople, John Roberts, Stephen Berry and Kenneth Pitzer. In the end, though, all major scientific progress arises from “gap jumping”, to quote Derek Barton. To do this, however, one must recognize the gap and then want to jump. Pioneers do both.

Science is dispassionate in its aims and international in its scope, and yet the activity of scientists is strongly influenced, even limited, by society. Take, for instance, Roald Hoffmann and Kenichi Fukui. These chemists shared the 1981 Nobel Prize in Chemistry for their theories, developed independently, concerning the course of

chemical reactions. Hoffmann would clearly like to convey a broader message to society through his poems, films and general writing, but recognizes his limitations when he concedes that, in the United States, scientists and their achievements are generally ignored.

Fukui, in contrast, is uncomfortable about communicating with the lay public on scientific matters, but is inundated with requests to do so. Is his austerity, quite typical of Asian cultures, a result of society's admiration or does it actually accentuate it? Again, is Hoffmann's obvious enthusiasm a response to the general lack of interest in science among the American public, or does a proactive stance, quite common in the West, induce the general apathy? The truth probably lies somewhere in between, but it would still be interesting to record the reactions of Nobel laureates from the United States and Europe to the near-hysterical adulation they receive in Asian settings, almost as a matter of course.

Another theme that emerges is that the most successful chemists appear to be able to change their research interests effortlessly, in some cases many times over, during their careers. Sherwood Rowland refers to the yawning gap between being “in the groove” and “in the rut”. Philip Eaton, who synthesized cubane in the 1960s using a stepwise route, is sure that synthesizing the similar but far more complex buckminsterfullerene using a similar approach today would be a waste of time. Fukui switched completely from experiment to theory.

Many of those interviewed also seem to

be concerned with the general shortcomings of the chemical community. Olah feels that chemists just don't think about the broader picture; he also admits that they are not the most interesting of people — surely, there is a connection. Hoffmann states blandly that any piece of junk can be published somewhere, and that even in the *Journal of the American Chemical Society*, the acceptance rate is around 60% for full papers. These and many other comments need to be read carefully and assimilated, especially by newcomers to the subject.

With the rapidly changing research scene, one is almost wistful about the past — when Djerassi exalts Robert Woodward and Robert Robinson as generalists, when one compares Elion's gentle and thoughtful approach to drug design with today's high-throughput screening procedures, when one savours the complete picture of marine natural products obtained from Scheuer's work, and when Hoffmann laments the lack of teaching content in a research paper, one feels that perhaps the golden age of classical chemistry is over.

Research is and will always be exciting, but the conversations in this book encapsulate a time that is past, and leave the reader with a comforting glow. The main protagonists have told their tales, and the editor has conducted his interviews with sympathy and collected his material with care. For this, he is to be commended. His book will be enjoyed by chemists and non-chemists alike. ■

Gautam R. Desiraju is in the School of Chemistry, University of Hyderabad, Hyderabad 500 046, India.

Putting science in its place

The Architecture of Science

edited by Peter Galison & Emily Thompson
MIT Press: 1999. 576 pp. \$65, £43.50

David N. Livingstone

Scientific knowledge is made in many different places; does it really matter where? To put it another way, can the location of scientific endeavour affect the conduct of science and, even more importantly, its content? The contributors to the present collection evidently think that the answer to these questions is an emphatic 'yes'.

On the surface at least, this is a remarkably counter-intuitive claim. Of all the human projects devoted to laying aside prejudices, and to putting in place mechanisms to guarantee objectivity, has science not been the most assiduous in executing its ideals?

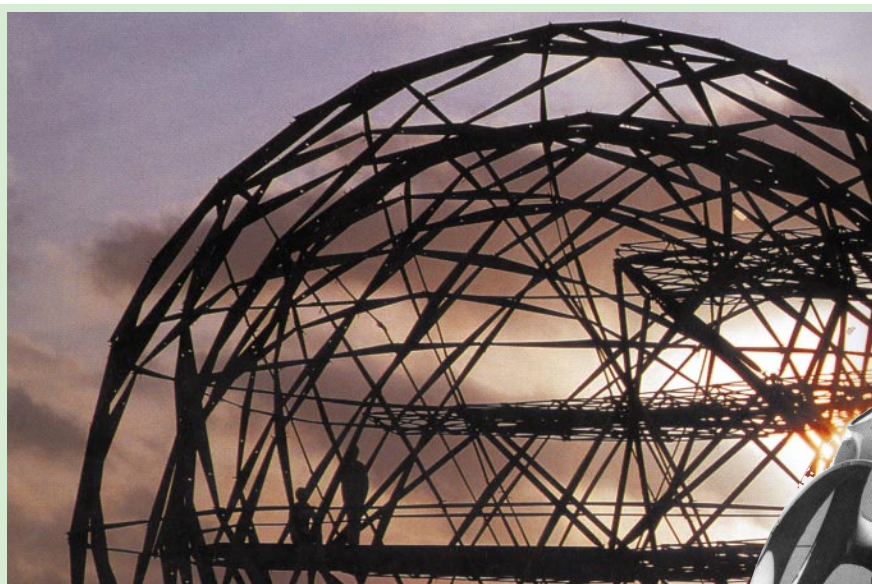
And yet science has been practised at a vast array of sites, each with different physical, acoustic and olfactory qualities: the alchemist's workshop with its roaring furnace and smelly, noisy stills; the wide-open, airy spaces of the field; the fusty alcoves of the museum; the antiseptic hospital. Even to express things in this way, of course, is to run the risk of caricature. Laboratories, gardens, observatories, hospitals and so on all come in a wide variety of sizes and configurations. But these stereotypes can convey something of the remarkable array of knowledge-producing scientific arenas.

Any attempt to come to terms with the spaces of scientific endeavour is plainly a multi-faceted project. And the essays in this collection focus on one key aspect of the task: the connections between science and architecture. The entire volume is concerned with elucidating the relationships between the buildings of science and the building of scientific knowledge.

Temporally, these essays, by academics and practitioners, take us from early modern European museums and chemical houses to twentieth-century molecular biology laboratories and the post-modern hospital. Conceptually the range is just as great, dealing with the ways in which the arrangement of scientific space has managed the tricky relationships between secrecy and openness, concealment and display; with the role architecture plays in shaping individual and group identity; and with the prevalence of physiological and mechanical metaphors (such as circulation and compression) in architectural thought. More specifically, the links between scientists and architects in the construction of the Lewis Thomas Laboratory for Molecular Biology at Princeton is the subject of several chapters.

Like most multi-authored works, this book lacks a single, coherent line of argument. Some of the essays consist of the autobiographical reflections of individuals directly involved in particular building projects; others are normative arguments about the kind of relations that should obtain between science and architecture; others are historical interrogations of how the shape of buildings influences the shape of science.

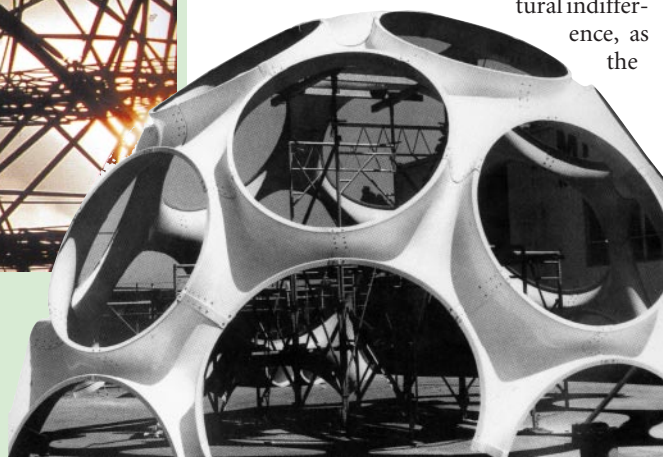
But the crucial issue, in my view, is whether (and if so, how) the cognitive content of science is influenced by its setting. Building arrangements have a bearing on the social relations that can take place among the scientists inhabiting these spaces. But can the architectural spaces themselves condition the knowledge that is produced? Whether this question can be answered without succumbing to either architectural determinism or architectural indifference, as the



Architectural adventures

A model of the Autonomous House (above) and Richard Buckminster Fuller's 'Fly's eye dome', from *Norman Foster: A Global Architecture* by

Martin Pawley (Thames & Hudson / Universe Publishing, £14.95/\$25).



editors put it, is of paramount importance.

Several essays bear directly on this issue. Consider, first, the topic of who makes scientific knowledge. The rhetoric of 'openness', so dominant in the early days of the scientific enterprise, was actually compromised by a number of strategic exclusions that did much to shape the nature of the endeavour: women, for example, were denied access to some sites of knowledge and, when they did engage in the pursuit of natural historical knowledge, had to do so in very different spheres. Again, early laboratories very carefully managed their thresholds in order to ensure that only the 'right' visitors were allowed access. The acquisition of scientific knowledge was thus part of a social process that had its own cultural topography.

But issues of 'access' are not the only way in which place and space have influenced scientific claims. Spatial arrangement has also been important. The ways in which nineteenth-century museums displayed their artefacts, for example, expressed different views about the nature and significance of the very objects themselves. Should certain items be displayed side-by-side or far apart? Should a specimen's site of discovery take precedence over its place in some taxonomic scheme when being presented for public scrutiny? The ways in which such questions were resolved disclosed how cognitive claims and spatial arrangements were mutually reinforcing. In such circumstances, museum space became a contested map of scientific judgement. In the case of anthropological museums, the physical layout of the exhibits conjugated differences between anthropological leaders on the very nature of their projects. Just how human history was displayed articulated different ways of reading the story of the species.

Hospitals also reveal intimate links between architectural configuration and claims to knowledge. In the early nineteenth century, hospitals were built to give expression to the belief that patients were in need of moral as much as medical help; they were intended to instil virtue as well as to restore health. Accordingly, the internal structure of the hospital was designed to impose order and control on the chaos of suffering and disease. Later, the advent of the multi-storey hospital mirrored the shift from bad-air theory to the germ theory of disease; the isolation presumed necessary for the former was no longer architecturally relevant in an age of sterilization. More recently, images of the mall or hotel have been increasingly favoured as the appropriate trope for the post-modern hospital.

Whatever its drawbacks, then, *The Architecture of Science* is a most welcome volume. Readers will never again look at scientific architecture with the same eyes. ■

David N. Livingstone is in the School of Geography, The Queen's University of Belfast, Belfast BT7 1NN, UK.

Science in culture

Prints and imprints

Chris Drury's "Journeys on Paper"

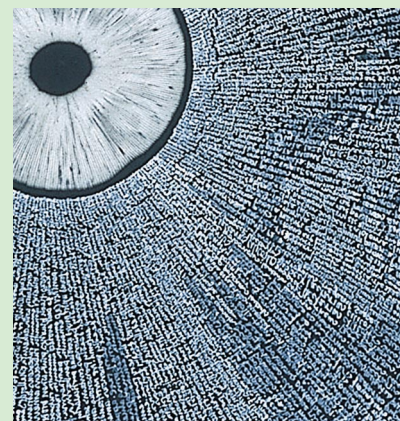
Martin Kemp

Some scientific observers of nature seem naturally drawn to complex phenomena, reaching out to grasp the elusive patterns underlying such fluctuating systems as populations of predators in relation to prey, or the beguiling chaos of fluids in turbulent motion. Others are attracted to the potential certainties of 'mathematical' engineering, in which the goal is to define the smallest functional units as components in the reconstruction of effects from causes. Temperament is clearly a powerful factor in determining who chooses which path.

Artists who aspire to reconstitute nature in their work — without necessarily imitating natural appearance — also tend to gravitate towards one of these two poles. Among the British predecessors of Chris Drury as students of landscape, Ben Nicholson's geometricizing reliefs and drawings, undertaken in St Ives between the two world wars, leant towards the mathematical pole, while Ivon Hitchen's contemporary oil paintings exploited free sweeps of overlaid paint to evoke the elusive contingencies of light, colour, atmosphere, reflection and motion in the moist English landscape.

Chris Drury's prime interest is the complexity of natural forces in action. Yet, like a number of recent artists who involve themselves in process rather than direct portrayal, he is drawn to the way in which the inherent structures in dynamic systems result in orders to which we can instinctively respond — even without benefit of the new mathematics of complexity.

Drury has worked extensively in nature itself, using natural materials to construct land sculpture and 'cloud chambers' (stone 'hives', which enclose camera obscura images of moving skies). His works on paper — or rather using paper as a surface to be manipulated — range widely across phenomena in which he senses patterns of affinity. Swirling folds in driftwood trunks of redwood are reminiscent of vortex configurations in a cross-section of tissues in the human heart. Caps of different mushrooms deposit their spore prints in a minute tracery of



radiating geometry that is at once regular and infinitely variable.

Yet there is something more at work than 'nature art'. Maps, those most conventional plottings of the surface of the Earth, are interwoven, basket-wise. A map of the Ladakh desert, for instance, is interwoven, strip by strip, with paper rubbed with desert earth to form a shallow bowl, which is in turn recessed within a rubbing from a prayer stone encountered en route. Maps are peppered with words. The nucleus of a spore print is typically surrounded by minutely inscribed names, phrases and clauses in a radial pattern that marvellously echoes the deposits from the interstices of the gills. In *Poison Pie* (pictured) the white spore print of *Amanita muscaria* (Fly Agaric) extends in an aureole of text that chants the names of poisonous fungi — *Amanita phalloides* (Death Cap), *Amanita virosa* (Destroying Angel), *Russula emetica* (The Sickener), *Coprinus atramentarius* (Ink Cap), *Hebeloma crustuliniforme* (Poison Pie itself), and so on — in concert with cool accounts of their identification and toxicity.

Drury patiently interweaves natural images and mental imprints in a constant give and take between the business of observation, acts of naming and recording, means of visual plotting, processes of classifying, evocative associations, inscribed memories, and the kinds of spiritual strivings that have accompanied so many cultures in their quest to become one with nature. As Marina Wallace says in her catalogue essay, "Drury accesses the scientific classifications and attempts to turn them into almost mythological narratives, using the element of repetition to accompany the visual marks". The result is an endlessly suggestive immersion in our visual and conceptual relationship with nature, in the orders we can discern and the contingencies in which we find such human delight. ■

Martin Kemp is in the Department of the History of Art, University of Oxford, 59 George Street, Oxford OX1 2BE, UK.

Chris Drury's "Journeys on Paper" are on view at the Stephen Lacey Gallery, One Crawford Passage, London EC1 3DP, until 7 July.



The microscope's coat of arms

...or, the sting of the bee and the moons of Jupiter.

Giovanni F. Bignami

Biologists and watchers of things small usually credit Anton van Leeuwenhoek (b. 1632), Marcello Malpighi (b. 1628) and Robert Hooke (b. 1635) as the first to put the microscope to scientific use, starting around the 1660s. But more than 30 years earlier, the school of Galileo, and in particular Francesco Stelluti, used the instrument to study the anatomy of insects, especially the bee. The beautiful drawings of Stelluti's *Melissographia* present the first micro-anatomical details of the bee, and are dedicated to Pope Urban VIII, in 1625.

Why the bee? One of Italy's oldest and noblest families, the Barberinis — who hail from the part of Tuscany that bears their name — date back to the time of Dante. Their original coat of arms featured three horseflies on a sky-blue background, but in the early 1600s, Maffeo Cardinal Barberini decided to introduce a touch of elegance, and the horseflies became honey bees, both sweeter and more socially acceptable.

The cardinal's motives became clear on 6 August 1623, when he was elected Pope Urban VIII. Now in a papal family emblem, bees became particularly important to the many seeking the Pope's favours. Gianlorenzo Bernini himself, the great baroque architect of the interior and colonnades of St Peter's in Rome, sculpted on his splendid Tritone fountain the Papal coat of arms, with bees so realistic that even today they still seem to be about to fly away. And the swarm of poets and literati around Pope Urban were themselves described as bees (*Apes Urbanae*) by the courtier Leone Allacci.

In the delicate relationship between church and science, Urban at first seemed to be well-disposed towards Galileo, as well as to Prince Cesi and his eclectic and elitist Accademia dei Lincei, the oldest scientific society in the world. What better moment, then, for the Lincei to study the detailed anatomy of the bee, and of course to dedicate their work to the Pope — especially as a new instrument had just started to become a scientific tool. Lens



Creating a buzz: Stelluti's detailed drawings of bees represent one of the first uses of the microscope.

arrangements for magnification had been around since about 1590, but the Lincei were the first to put the instrument to serious scientific use, and to record and publish the results.

In Galileo's times, however, the optical instruments for looking at distant things or at very small things did not yet have accepted, universal names. Galileo called the instrument more seriously known as *specillum* or *tubus bilens* (a pipe with two lenses), his *occhiale*. Prince Cesi called it, aptly, *uranoscopia*. Around 1611, an agreement was reached on *telescopio*. Galileo's instrument was tiny compared with today's: it went just two magnitudes below the power of the naked eye. But with it he brought about a revolution in astronomy, cosmology and philosophy, starting with the moons of Jupiter, the revolutions of which were enough to shatter the old sky paradigms.

As for the other instrument, Galileo referred to it as "the small glass for spying things up close", but the term *enghiscopio* had also been proposed, as well as the contrived *ponoscopia*. In the end, it was Stelluti who, around 1625, introduced *microscopio*, a name that, luckily, stuck. Thus, the two great optical instruments of the scientific revolution were put to use, and also received their current names, within less than two decades, from 1611 to 1625.

To return to bees, the similarity of the frontispiece of Stelluti's *Melissographia* to the Barberini coat of arms is obvious. But here the three bees are shown in dorsal, ventral and lateral projections, and anatomical details are also shown separately, at the base of the image. The drawing renders a wealth of details in precise and vivid terms, and the Pope, who was very pleased, confirmed his benevolence to the academy — if not to Galileo, above whom the clouds of the Inquisition were gathering.

A few years later, Stelluti published his detailed and systematic *Description of the Bee*, including its body, its parts and organs, down to its "faceted eyes covered in fur" and its tongue "surrounded by four tonguelets". It represents probably the first — and one of the most stylish and complete — descriptions of what was actually observed through a microscope. Its method and its (Italian, rather than Latin) prose are reminiscent of Galileo's descriptions of the heavenly bodies he saw through his telescope. Stelluti's microscope was of modest magnification, not much better than the resolution of the naked eye. And yet it showed, for example, the bee's sting, never before seen clearly. It should thus rank, in the history of science, on a par with the moons of Jupiter. ■

Giovanni F. Bignami is at the Italian Space Agency, Viale Liegi 26, Rome 00198, Italy.

Bees became particularly important to the many seeking the Pope's favours.

Tuberculosis bacteria join UN

WHO proposes to include disinfectant under the Geneva Convention.

Joan Slonczewski

A milestone in microbiology was passed today (29 June) when *Mycobacterium tuberculosis* ssp. *cyberneticum* was voted full membership of the United Nations (UN).

Seena Gonzalez, director of the World Health Organization (WHO), reflected on the significance of the UN's acceptance of the first cybermicrobe, despite the notoriously murderous history of its ancestral species. "It's probably true that bacteria invented mass homicide," she concedes, "but then, second-millennial humans perfected the art. If Stalin joined the UN, why not TB?"

The evolution of microscopic intelligence was predicted at the turn of the millennium by Beowulf Schumacher, a physics professor at a small college in rural North America surrounded by cows carrying *Escherichia coli*. Schumacher predicted the development of nanocomputers with computational elements on an atomic scale, based on principles of cellular automata.

The first nanobots — primitive by today's standards — were used to navigate the human bloodstream, where they cleaned up arterial plaque, produced insulin for diabetics, detected precancerous cells, and modulated neurotransmitters to correct mental disorders. But initially, the survival of nanobots *in vivo* was poor, and their failure caused serious circulatory problems.

Then, in 2441, investigators at the Howard Hughes Martian Microbial Institute hit upon the idea of building computational macromolecules into the genomes of pathogens known for their ability to infiltrate the human system. After all, the use of pathogens such as adenovirus and HIV as recombinant vectors was ancient history. Why not build supercomputers into some of humankind's most successful pathogens?

M. tuberculosis was a prime candidate — it inhabits the human lungs for decades, in the ideal position to seek and destroy any pulmonary cells transformed by inhaled carcinogens. Tobacco companies poured billions of dollars into developing cybernetically enhanced, cancer-sniffing TB.

What no one anticipated was that the enhanced bacteria, like so many macroscale robotic entities in the past century, would develop self-awareness and discover a true brotherly love of their human hosts. "Let's face it," says a TB spokesperson, "we never really wanted to kill humans anyway. Our

ancestors inhabited humans peacefully most of the time, for hundreds of generations. Occasionally we messed up and trashed our environment — but how many human nations haven't?"

TB's acceptance has been met with some controversy in the bacterial community. In particular, some isolates of *E. coli* K-12 feel miffed that their own request for membership was not granted first. "*E. coli* has always been the molecular biologist's best friend," K-12 points out. "Why weren't we accepted first? We didn't even get our genome sequenced first. Life is unfair."

K-12 also noted that *E. coli* and other human commensals have suffered centuries of abuse from their hosts, as medical and research institutions conducted mass slaughter of harmless bacteria through the indiscriminate application of antibiotics. The North American National Institutes of Health has recently signed a treaty with several cybermicrobial species, in which the institute researchers promised to respect the independence and survival rights of cybermicrobial colonies. "Thank goodness the sun finally set upon their colonial empire," K-12 observes pointedly.

On the positive side, the National Science Foundation (NSF) was applauded for its more benevolent approach over the centuries, even declining to support medically oriented antimicrobial research. "NSF's curiosity-driven researchers have created wonderful new strains of curious microbes," comments veteran panellist Meheret Beck. "The grant proposals submitted by these microbes often get rated as 'Outstanding'."

One such outstanding project is that of cyber-*Helicobacter*. The gastric bacteria propose to engineer themselves to convert highly caloric foods into molecules that pass undigested through the intestinal tract, thus helping their human hosts avoid excessive weight gain. "Of course, digestive microbes have long helped animal hosts accomplish the opposite," notes Beck.

Biomedical researchers remind us, however, that not all microbes have given up their war on humans — many deadly species remain unreconstructed. The so-called Andromeda strain, for example, is still under the sway of an unstable dictator who vacillates between homicidal frenzy and paranoid isolation.

Nevertheless, the extraordinary flowering of democratic civilization among cybermicrobes has won the admiration of many human nations, even those who themselves still decline UN membership. As Swiss spokesperson Ursula Friedli observes: "Microbes, unlike their metazoan relatives, have always eschewed centralized organization in favour of more democratic cooperative structures such as biofilms. We Swiss can relate to that." Friedli, however, denies rumours that the cybermicrobes' example will finally convince Switzerland to join the UN. "Maybe after the Alzheimer prion joins, we'll consider it," she admits. "But for now, persecuted microbes seeking refuge from WHO can apply for asylum in our neutral country."

Joan Slonczewski is a microbiologist at Kenyon College, Gambier, Ohio. Her novels include *The Children Star* and *Brain Plague* (<http://www2.kenyon.edu/depts/biology/slonc/slonc.htm>).



Critical time for fluid dynamos

Andy Jackson

Einstein once cited the origin of the Earth's magnetic field as one of the fundamental unsolved problems of physics. Although there has been progress on paper, experimental models have been elusive until now.

The Earth's magnetic field originates in the liquid metallic core of the Earth, where it is sustained by complicated fluid motions. Despite some progress on the underlying theory, attempts to recreate this system in the laboratory had, until recently, failed to surmount the enormous technical difficulties involved. But many years of effort have now come to fruition, with reports of successful experiments at Riga in Latvia¹ (from a Latvian–German collaboration) and at Karlsruhe in Germany². These experiments finally create the right conditions under which an isolated fluid can grow and maintain a magnetic field. The fluid systems are similar, though far from identical, to the Earth's own, but the experimental results are underpinned by calculation, and pave the way for more realistic simulations.

Inspired by Sir Joseph Larmor's explanation³ for the magnetic field of the Sun, geophysicists first proposed in 1919 that the Earth's field is generated by a 'self-exciting dynamo' in the planet's core. In this model, motion of the electrically conducting core through an already existing magnetic field induces electrical currents within the core, in the same way that a bicycle dynamo (or generator) generates a current. The magnetic field associated with these currents amplifies the original field, ensuring that it does not decay with time. A weak, interplanetary magnetic field could provide the original source field for this process.

It is easy to design contrived mechanical systems that perform this positive feedback, though they are rather different from the Earth (the quintessential example being the Bullard disc dynamo⁴). But to mimic the Earth's dynamo we need to use a homogeneous material. The first dynamo made of homogeneous material was the Lowes–Wilkinson dynamo^{5,6}, which consisted of two solid cylinders spinning at right angles to each other within a block of metal (Fig. 1a). The cylinders represent swirling eddies in an otherwise static background, and show that a simple homogeneous material can work as a self-exciting (or self-sustaining) dynamo.

For a model to be more realistic it should involve conducting fluids, but it is much harder to show that they work as self-exciting dynamos. Early work on fluid-based models suffered several setbacks. For example, the Earth's field is largely axisymmetric, and the proof⁷ that a fluid sphere cannot maintain

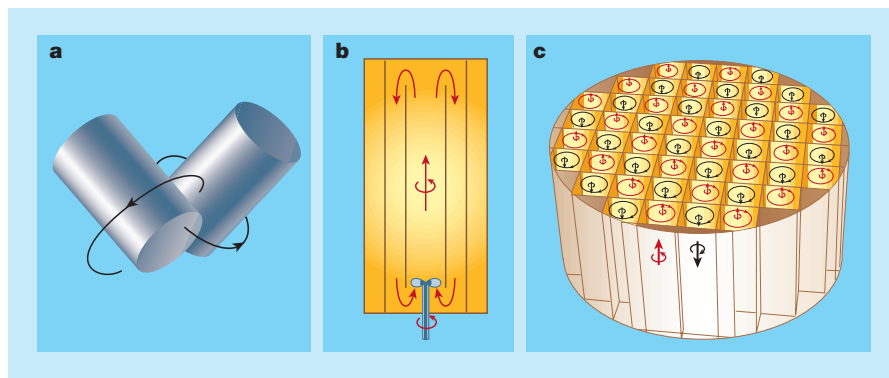


Figure 1 Helping dynamos go critical in the lab. To recreate the Earth's molten outer core, researchers need to build a dynamo made from a homogeneous fluid. a, The Lowes–Wilkinson dynamo; the rotating cylinders are embedded in a block of metal. Although it was solid, it was the first homogeneous dynamo. b, The Riga dynamo¹ (side view); the helical flow in the centre returns through a surrounding tube, which is encased in static fluid. c, The Karlsruhe dynamo²; each cell has a vortex motion of opposite sense, shown by the red and black arrows. The dynamo will work only if it can generate new magnetic fields more quickly than they decay — this is measured by the magnetic Reynolds number. In these three dynamos the decay is caused by the finite electrical resistance of the metals used. When the Reynolds number exceeds a critical value the material will generate a self-sustaining magnetic field.

an axisymmetric magnetic field initially threatened to kill off the theory. Indeed it took almost 40 years for the theory behind self-sustaining fluid dynamos to finally be proved. The advent of computers has led to several successful solutions of the complex equations underlying the theory, albeit under modelling conditions that are rather different from the Earth's (see ref. 8 for a review). But there remained the nagging feeling that one should be able to demonstrate the field-generation mechanism in the laboratory, not only as a verification of the theory, but also as a complementary line of research to the numerical models. That has now been achieved^{1,2}.

To build a fluid dynamo one must use a very good conductor, usually liquid sodium, with all its incumbent dangers. Sodium is much more reactive as a liquid than as a solid; it will burn easily in air, and its reaction with water can be explosive. The dynamo will work only if the motion of the magnetic field, and generation of current, are fast compared with the rate at which currents die away because of the finite electrical resistance — this is measured by the so-called magnetic Reynolds number, which must be reasonably large. When this number exceeds a critical value, equilibrium (or steady-state) field generation can occur, in the same way that happens in the Earth.

The Earth achieves a large magnetic Reynolds number — despite the ponderous fluid motions in the core (of the order of half a millimetre per second) — because of its enormous dimensions; in the laboratory, a much faster flow speed is required. Fluid motion in the Earth's core is driven by convection, which transports heat from the interior out to the mantle, but this method of generating fluid motion is too inefficient for the laboratory dynamos, so propellers and pumps are used instead.

The latest experiments at Riga¹ and Karlsruhe² both use spiral flow to stir the molten metal, but the two systems differ in their geometries. In the Riga experiment a single vortex in the centre flows back through a surrounding tube, which is itself enclosed by fluid at rest (Fig. 1b). The Karlsruhe group use a circular 'honeycomb' structure more than a metre across, consisting of 52 tubes within which fluid circulates in a vortex. The flow in adjacent tubes rotates in opposite directions, as well as flowing in different vertical directions (Fig. 1c). Both experiments rely on the twisting motion of the fluid flow to amplify the magnetic field. The Riga experiment requires flow rates as large as half a cubic metre per second, whereas the Karlsruhe experiment, by dint of its larger number of helices, can

manage with flow rates of roughly two cubic metres per minute.

Late last year, both groups succeeded in growing magnetic fields from tiny 'seed' fields. Both experiments are backed by extensive calculations, and the authors have compared their results with those predictions, finding reasonably good agreement. These positive results bode well for the burgeoning number of experiments of this type around the world: there are at least six groups in France, Germany and the United States, each looking at different aspects of fluid dynamos (reviewed by ref. 9).

Given the difficult nature of the physical equations underlying dynamo theory, theorists use a simplification known as 'mean-field dynamos', in which values for the very small-scale fluid motions are represented in a statistical sense rather than represented explicitly. Although this is a plausible mathematical tool, it has never been tested directly until now. The results from the new experiments are encouraging. The Earth's core is expected to be turbulent, and the fluid flows in both experiments are turbulent. The two experimental set-ups complement each other, in that the turbulence in the Karlsruhe dynamo does not play a crucial role in deter-

mining the magnetic field, whereas it may be significant in the Riga experiment. In view of the fact that turbulent systems are so difficult to model numerically, future experiments will test our understanding of this process.

These experimental dynamos also operate in a regime that so far has been inaccessible to numerical simulations. In this regime viscosity plays a very small role in comparison to the resistive decay of the magnetic fields. Further experiments will hopefully allow us to probe this area. Given dynamo theory's unpromising start, the past decade has been remarkable: experimental dynamos have emerged at the same time that large numerical simulations have reached some maturity. There is now hope for real progress. ■

Andy Jackson is at the Department of Earth Sciences, University of Leeds, Leeds LS2 9JT, UK.
e-mail: jackson@earth.leeds.ac.uk

1. Gaillitis, A. *et al.* *Phys. Rev. Lett.* **84**, 4365–4368 (2000).
2. Müller, U. & Stieglitz, R. *Workshop on Homogeneous Dynamos* Forschungszentrum Karlsruhe, 20–22 March 2000.
3. Larmor, J. *Br. Assoc. Adv. Sci. Rep.* 159–160 (1919).
4. Bullard, E. C. *Proc. Camb. Phil. Soc.* **51**, 744–760 (1955).
5. Lowes, F. J. & Wilkinson, I. *Nature* **198**, 1158–1160 (1963).
6. Lowes, F. J. & Wilkinson, I. *Nature* **219**, 717–718 (1968).
7. Cowling, T. G. *Mon. Not. R. Astron. Soc.* **94**, 39–48 (1934).
8. Busse, F. H. *Annu. Rev. Fluid Mech.* **32**, 383–408 (2000).
9. Tilgner, A. *Phys. Earth Planet. Int.* **117**, 171–177 (2000).

Genetics

Targeting sheep

Milind Suraokar and Allan Bradley

Over the past few thousand years, the physical characteristics of sheep and their underlying genetic make-up have been slowly altered by selective breeding from animals with desirable characteristics. The result is that sheep are now mobile protein factories that can survive in harsh environments and still provide abundant quantities of milk, meat and wool. But although our ancestors succeeded in significantly altering this species, selective breeding is limited by the pool of desirable variants in the sheep population. Transgenic technology, developed over the past 15 years, has begun to address this problem. But, although this approach can be used to add new genetic information, it is not possible to specifically modify the genes of the species itself by this route. McCreath and colleagues, writing on page 1066 of this issue¹, have finally overcome this problem by making use of the now familiar technique of cloning.

Transgenic technology has been widely applied to livestock species because it is technically quite simple, requiring only the injection of 'naked' DNA — unencumbered by its protein partners — into the nucleus of a fertilized egg². Genes encoding useful human

proteins such as $\alpha 1$ -antitrypsin (an enzyme that is mutated in the disorder familial emphysema) have been inserted into sheep eggs in this way, resulting in transgenic sheep that express the human protein in their milk³. But despite the ease with which transgenic animals can be generated, this technology has limitations, because the injected DNA integrates randomly into the sheep genome. Often the injected DNA does not land in a site in the genome that allows the foreign gene (transgene) to be expressed in the desired tissue or at the appropriate level. Moreover, the sheep's endogenous genes cannot be specifically altered using this technique.

By contrast, researchers have been able to modify endogenous genes in laboratory mice for over ten years⁴. This type of genetic manipulation became possible only because cells known as embryonic stem (ES) cells could be isolated from mouse embryos⁵. Specific mutations can be generated, or specific gene sequences altered, in cultured ES cells; rare variants with the desired genetic change can then be selected. These modified cells retain their developmental potential and, when inserted into a developing embryo, can contribute to all of its tissues, including the

germ cells (egg and sperm). So the changes selected in cultured cells can be transferred into the genetic repository of the animal.

In the mouse, ES cells have provided the essential link between selection schemes in cell culture and the whole organism. The usefulness of this system for producing genetic changes has been widely exploited⁶. For more than a decade, researchers have searched for ES cells from a multitude of other species, including livestock. But, although some cell lines have been reported, none has been able to pass the crucial test — that of contributing to germ cells.

So the hope that livestock could be genetically manipulated like mice languished until a few years ago, when it was revived by the cloning of livestock species⁷. This technique, later extended to the use of nuclei obtained from cultured adult cells⁸, has circumvented the need to isolate the elusive ES cell from species other than mice, and is a powerful way of transferring transgenes to livestock via the culture dish^{9,10}. However, all the genetic changes established so far by cloning could have been achieved using conventional transgenic technology. Until now, no one had shown that it would be possible to specifically modify endogenous genes by cloning.

The success reported by McCreath *et al.*¹ hinges in part on strategies designed to overcome some of the perceived problems of performing gene targeting in fibroblasts — the cell type of choice for cloning — rather than ES cells. For instance, they selected the *COLIA1* genetic locus as the endogenous sequence to be altered, or 'targeted', because this gene is highly expressed in fibroblasts. The authors were concerned that targeting efficiencies in fibroblasts might be much lower than the 5–20% typically observed in mouse ES cells (these percentages are ratios of desired to random integration events). The high level of expression of *COLIA1* allowed McCreath *et al.* to use a 'promoter-trap' strategy (similar to that used in early attempts to target genes in ES cells¹¹) to increase their chances of detecting targeting events at the *COLIA1* locus.

In one set of experiments, the authors inserted a *neo* cassette — a stretch of DNA encoding a protein that allows cultured cells to grow in the drug G418 — downstream of the *COLIA1* gene in sheep fibroblasts. In a second set of experiments, the targeting sequence also included a transgene downstream of the *neo* cassette (Fig. 1). This transgene encoded human $\alpha 1$ -antitrypsin, as well as a sequence to ensure that the transgene would be switched on only in the mammary gland tissue of lactating ewes. The targeting efficiency with both vectors was similar to that obtained using ES cells.

Fibroblasts cannot proliferate for long *in vitro*, and acquire chromosomal changes quite quickly. This meant that the authors had to achieve targeting and identify targeted

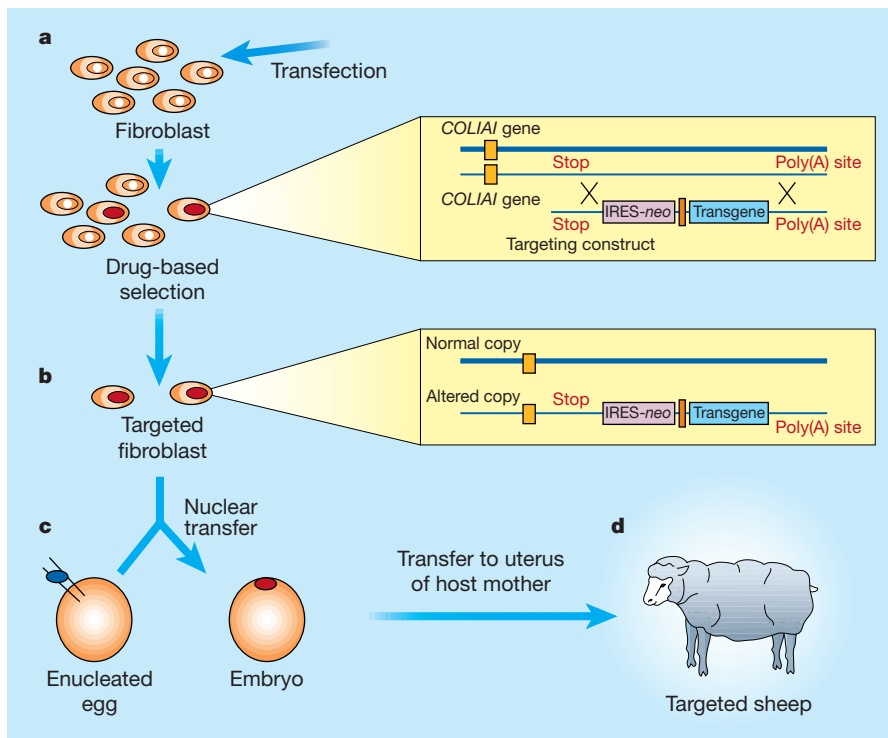


Figure 1 Modifying specific genes in sheep. McCreath *et al.*¹ have created sheep with a human gene inserted into a particular genetic locus. **a**, The targeting vector — which carried the neomycin-resistance gene (*neo*) and a foreign gene (transgene) encoding human $\alpha 1$ -antitrypsin — was inserted into sheep fibroblast cells. The internal ribosome-entry site (IRES) allowed the messenger RNA produced from the transgene to be translated into protein. The transgene and selection cassette were incorporated between the translational stop site and the polyadenylation (poly(A)) site of the sheep *COL1A1* locus. Transfected cells expressed the *neo* gene and so could be grown in the drug G418 (a neomycin analogue). **b**, The DNA of such cells was analysed by the polymerase chain reaction to verify that the *COL1A1* locus was targeted. **c**, These cells were used for nuclear transfer. Nuclei were removed from sheep eggs, which were then fused to single transfected cells. The resulting embryos were transferred to foster ewes. **d**, Some of the implanted eggs developed into healthy lambs expressing $\alpha 1$ -antitrypsin in their milk.

clones rapidly. So they transferred the targeting vectors to the fibroblasts early in their growth history, and identified targeted cells by their ability to grow in G418 and, later, by molecular techniques (the perennially useful polymerase chain reaction).

The next step was to fuse the transfected fibroblasts with an egg from which the nucleus had been removed, and to allow the egg to develop into an embryo (Fig. 1) — the protocol developed for cloning. Using this strategy, the authors established two types of targeted loci in sheep. When the targeting sequence encoding $\alpha 1$ -antitrypsin was used, high levels of this protein were found in the sheep's milk. So the *COL1A1* locus might be a good site in the genome to place mammary-gland-specific transgenes.

This technique should provide a general way to produce specific genetic changes in several mammalian species. But the opportunities extend far beyond the optimal positioning of foreign genes. Mouse models of several human diseases — such as cystic fibrosis¹² — are not always precise because of differences between the two species. Perhaps the application of McCreath *et al.*'s technique

to more closely related mammals might be of use here. In agriculture, undesirable genes could be removed, such as *PrP*, in this case producing flocks of sheep that are resistant to scrapie¹³. We are clearly at the dawn of a new era in mammalian genetic technology.

Milind Suraokar and Allan Bradley are in the Department of Molecular and Human Genetics, Howard Hughes Medical Institute, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030-3498, USA.

e-mails: abradley@bcm.tmc.edu
ms690791@bcm.tmc.edu

1. McCreath, K. J. *et al.* *Nature* **405**, 1066–1069 (2000).
2. Hammer, R. E. *et al.* *Nature* **315**, 680 (1985).
3. Wright, G. *et al.* *Biotechnology* **9**, 830–834 (1991).
4. Bradley, A., Zheng, B. & Liu, P. *Int. J. Dev. Biol.* **42**, 943–950 (1998).
5. Evans, M. J. & Kaufman, M. H. *Nature* **292**, 154–156 (1981).
6. Ramirez-Solis, R., Liu, P. & Bradley, A. *Nature* **378**, 720–724 (1995).
7. Campbell, K. H. S., McWhir, J., Ritchie, W. A. & Wilmut, I. *Nature* **380**, 64–66 (1996).
8. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. S. *Nature* **385**, 810–813 (1997).
9. Schnieke, A. E. *et al.* *Science* **278**, 2130–2133 (1997).
10. Cibelli, J. B. *et al.* *Science* **280**, 1256–1258 (1998).
11. Schwartzberg, P. L., Robertson, E. J. & Goff, S. P. *Proc. Natl Acad. Sci. USA* **87**, 3210–3214 (1990).
12. Dorin, J. R. *et al.* *Nature* **359**, 211–215 (1992).
13. Prusiner, S. B. *Proc. Natl Acad. Sci. USA* **95**, 13363–13383 (1998).



100 YEARS AGO

There are, it is estimated, about 400 lepers in France. They are scattered about in Brittany, in the Pyrenees, on the shores of the Mediterranean, and in Paris, where they number 150. Among the lepers there are missionaries and nurses who have fallen victims to their devoted care of sufferers in other countries, and officials and soldiers who have contracted the disease in the colonies. An anti-leprosy committee has, says the *British Medical Journal*, recently been formed... for the care of the lepers in France and the prevention of the spread of the disease.

EXCELLENT results have been obtained by the French Government from experiments made with wireless telegraphy. The *Engineer* of June 15 says that the demonstrations showed that communication could be maintained, between ship and shore, to a distance of about sixty miles with comparative ease, only the height of the masts of the Government ship *Utile* preventing longer distances being attained. In consequence of these achievements the French Government have decided to equip their Mediterranean Squadron with the necessary apparatus. From *Nature* 28 June 1900.

50 YEARS AGO

There exists in the Cavendish Laboratory an extensive collection of apparatus used in the great era of classical physics. The start of modern physics is generally identified with J. J. Thomson's experiments on the electron in 1896, but a revolution of equal importance was taking place in the time of Maxwell and Rayleigh, the first two holders of the Cavendish chair. Physics was becoming established as an experimental science... and for the first time a home was being officially provided for it by the University, where students could do experiments and the staff could pursue research. It is rather startling to realize at what a late date the idea of an experimental laboratory took shape. The earliest record of anything of the kind appears to be the laboratory in a wine-cellar in one of the professors' houses at Glasgow, which Kelvin fitted up in 1846... Until that time the tradition throughout British universities had been that while the professor gave his lectures in natural philosophy in university lecture rooms, research was his private concern, to be carried out where he could. From *Nature* 1 July 1950.

Condensed-matter physics

Misbehaviour in metals

Philip B. Allen

A numerical study by Gunnarsson and Han¹ on page 1027 of this issue sheds new light on the electrical resistivity of complex metals. Resistivity is a basic property of materials, defined as electrical resistance per unit length of a wire of unit cross-sectional area. Standard measurements, such as the variation of resistivity with temperature, material purity and magnetic field, can tell us a lot about electrons in matter. In simple metals we can treat electrons as a gas, allowing the resistivity to be analysed in detail. But in more complicated metals, such as those studied by Gunnarsson and Han, the situation is more confusing.

In an ordinary gas, like the air we breathe, the uncharged molecules are well separated from each other, fly around at random thermal velocities (typically 1 km s^{-1}), and occasionally bump into each other. The mean distance between collisions is about 100 nanometres (100 times bigger than the mean distance to the next molecule). The particles are small and weakly interacting. By contrast, in a liquid the molecules continually bump into each other. In the 1870s, Ludwig Boltzmann invented a statistical theory that successfully describes the behaviour of gases and has since been adapted to describe a huge range of phenomena in physics and engineering. Unfortunately, it can't work for liquids, and no corresponding theory has ever been found².

In a simple metal, such as copper, there is one 'free' negatively charged electron for each positively charged copper ion in the crystal lattice. In 1904, Lorentz applied classical Boltzmann gas theory to these electrons (Fig. 1). On the face of it, the idea looks wrong, as electrons are neither dilute nor weakly interacting. They should collide after moving only a fraction of a nanometre, like molecules in a liquid. Nonetheless, gas theory does work for ordinary metals. Bloch later added quantum effects to show how electron waves diffract around the crystalline array of ions, not colliding until they encounter an impurity or a thermally displaced ion. But all these theories ignored interactions between electrons. In 1956, Landau introduced new ideas that led to our modern understanding of metals.

Through Landau's work we have come to recognize that, rather than having 'free electrons', a conducting metal actually has 'quasiparticle excitations' carrying the electrical current. These quasiparticles closely resemble electrons, and obey a modified gas theory. So electrons in copper behave like particles in a dilute gas. The electrical resis-

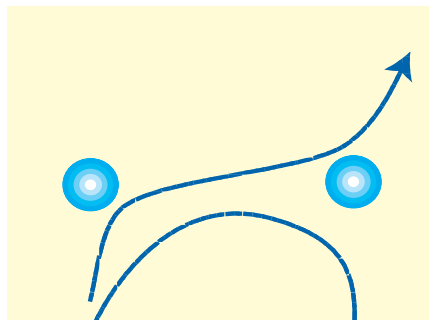


Figure 1 Modelling electrons as a gas. If one collision ends before the next begins (upper trajectory), classical Boltzmann gas theory applies. Otherwise (lower trajectory) the effect of the two collisions is not additive and gas theory fails. This is the problem with Boltzmann theory when a dilute gas is compressed into a dense liquid. In metals the quantum version of Boltzmann gas theory has more subtle problems leading to failure in complex metals, such as the metallic fullerenes studied by Gunnarsson and Han¹, and high-temperature superconductors.

tivity has been explained in fine detail, including the way it increases with temperature, as the distance between collisions reduces. In this semiclassical picture, the distance between collisions cannot be smaller than the lattice spacing, so the resistivity cannot be calculated reliably beyond a maximum temperature or impurity level.

The lack of a 'Boltzmann theory for liquids' is a big problem when it comes to interpreting resistivity in more complicated metals. For example, in high-temperature superconductors the resistivity continues to rise even as the distance between collisions gets too short to justify gas theory. Similar effects occur in superconducting metals made from 'buckyballs' (C_{60} molecules) doped with alkali metals³. There are suspicions that electronic behaviour in these exotic metals may be radically different from that in conventional metals.

Gunnarsson and Han¹ now model the resistivity of these 'metallic fullerenes' using quantum-mechanical π -electrons of the outermost shell of the C_{60} molecule, coupled to vibrations of the molecule that are also treated quantum mechanically. The electrons can hop from C_{60} to C_{60} , with partially disordered hopping because the C_{60} molecules are not fully in line with the crystal lattice. This model contains much of the physics expected to control the behaviour of the electrons, and is too complicated to solve except by a computer simulation. The result of the model is a plot of resistivity against

temperature that agrees with the experimental observations. The theory yields the same answer as a naive model in which there is no fundamental limit to the mean distance between collisions, at least for fullerenes.

What are the implications of Gunnarsson and Han's discovery? The Landau theory that underlies the modern theory of ordinary metals has a built-in circularity: one cannot prove that electrons in copper behave like particles in a gas, but one can show that this is a mathematically consistent possibility⁴. Other possibilities exist, the most dramatic being conventional superconductivity, in which attractive interactions between electrons destroy the quasiparticle gas phase of electrons at low temperatures, replacing it with the superconducting phase with zero resistance. The particles that fly around in a superconductor are then no longer electron-like quasiparticles, but become 'Bogoliubov quasiparticles'. Their electrical charge is altered from the usual value because the particles are a subtle mixture of negatively charged electron and positively charged 'hole'. These quasiparticles obey gas laws, but not the electronic version⁵.

A second alternative to the Landau theory is Anderson localization (Fig. 2). This is a theoretical possibility that arises when the electrons suffer very strong scattering from defects. As defect density increases, electrons evolve from free travelling waves into trapped wave packets that are unable to diffuse (localization). A third, more pedestrian possibility is paradoxically less well understood: as defect density or thermal ionic displacements increase, distances between collisions can reduce to the atomic-spacing level and resistivity can stop increasing and 'saturate' (Fig. 2). This saturation of resistivity is reminiscent of the evolution of a classical gas into a classical liquid, where gas theory no longer applies. Several other options (spin- and charge-density waves, for example) are also known.

Do the known options exhaust all the possibilities? Surely not. For example, a peculiar phase occurs in two-dimensional metals in strong magnetic fields, where electrons and magnetic flux can combine to form new com-

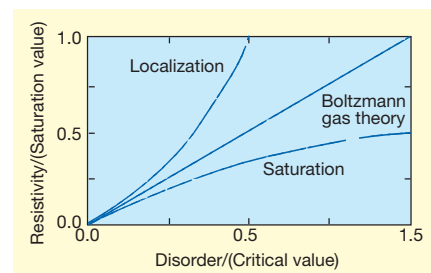


Figure 2 Alternative forms of electrical resistivity. The straight line shows the prediction of quantum gas theory, which works for simple metals. The other options (Anderson localization and 'saturation') might apply when materials contain a high density of defects or strong scattering from thermal disorder.

posite quasiparticles with fractional charge. The enigma of high-temperature superconductivity, which occurs in copper-oxide metals, has been driving the search for new theories. Above the temperature (around 100 K) where superconductivity disappears, the electrical resistivity of the copper oxides is highly unusual. The temperature dependence suggests some sort of quasiparticle gas, but the scattering is so strong that liquid behaviour should be seen⁶. This is often interpreted as evidence for a new exotic electron phase.

In their model for fullerenes, Gunnarsson and Han¹ find no indication that the electrons have formed such a phase. They also argue that the quantum Boltzmann gas picture continues to work reasonably well into the liquid-like regime, where it does not really apply.

So, for this model, it appears that unusual resistivity can be reconciled with ordinary behaviour. A similar theory for copper-oxide metals above their high superconducting transition temperature is needed. But this will be an even bigger challenge. Strong magnetic electron-electron interactions will need to be included in the calculation before the copper oxides can be adequately modelled. ■

Philip B. Allen is in the Department of Physics and Astronomy, State University of New York at Stony Brook, Stony Brook, New York 11794-3800, USA.
e-mail: philip.allen@sunysb.edu

1. Gunnarsson, O. & Han, J. E. *Nature* **405**, 1027–1030 (2000).
2. Peierls, R. E. *Surprises in Theoretical Physics* (Princeton Univ. Press, 1977).
3. Vareka, W. A. & Zettl, A. *Phys. Rev. Lett.* **72**, 4121–4124 (1994).
4. Shankar, R. *Rev. Mod. Phys.* **66**, 129–192 (1994).
5. Kivelson, S. A. & Rokhsar, D. S. *Phys. Rev. B* **41**, 11693–11696 (1990).
6. Emery, V. J. & Kivelson, S. A. *Phys. Rev. Lett.* **74**, 3253–3256 (1995).

Linguistics

Talking trees tell tales

Rebecca L. Cann

On page 1052 of this issue¹, Russell Gray and Fiona Jordan describe how they have used a tool from molecular taxonomy, the ‘parsimony’ method, to examine patterns of DNA sequence change over time and to probe the history of a group of languages. Parsimony is a central principle in the cladistic approach to evolutionary biology; it holds that the evolutionary tree requiring the fewest transformations of characters should be preferred. Words do not fossilize. Yet they leave evidence of their evolution in the populations that speak them, in much the same way that genes reveal the evolutionary history of the populations that transmit them.

Gray and Jordan’s subject of study is the spread of the Austronesian language family eastwards across the Pacific several thousand years ago. What makes their approach novel is the explicitly quantitative test of the hypothesis that the evolution of a material culture, as traced by archaeological artefacts and other evidence, can be mapped onto the evolution of a language family in ordered and nested steps.

Both comparative linguists and archaeologists are interested in the apparently rapid expansion of the first humans to reach remote Oceania, the region containing Pacific islands beyond the Bismarck Archipelago that includes traditional Polynesia along with parts of Micronesia and Melanesia (Fig. 1). Expansion began with a drive out of south-east Asia, southern China, Taiwan or Indonesia about 6,000 years ago. The exact geographical location of the founders of what would eventually become a unique, deep-ocean voyaging culture is not known for sure. But a linguistic analysis by Robert Blust²

favours a homeland in Taiwan, and Gray and Jordan’s analysis proceeded on that basis.

The earliest settlements in Fiji, western Polynesia and New Caledonia came less than 500 years after a push out of western Melanesia, and there is genetic evidence that Austronesian voyagers mixed along the way with coastal residents of Papua New Guinea and Vanuatu. Archaeological sites indicat-

ing colonization of islands beyond the Bismarck Archipelago are associated with the Lapita cultural complex, named after a characteristic type of pottery. Residents of these islands, presumed to be descendants of the original voyagers, speak languages grouped into the Austronesian family, which now contains about 1,200 distinct languages.

Parts of the sequence of initial colonization have been obscured by subsequent trading and warfare, along with depopulation and replacement in modern times resulting from disease epidemics. Also, gaps in the archaeological record, however short, may reflect either actual pauses in the speed of colonization or simply the lack of exploration of — and so evidence from — coastal sites that now lie under water.

It seems that Austronesian voyagers apparently stopped and rested for about a thousand years between the settlement of western Polynesia and the rest of remote Oceania. Blust suggested that this hiatus was due to a pause before the invention of the double-hulled sailing canoe that could cross the wider stretches of sea involved. Further, some archaeologists have long held that Polynesians did not acquire their full suite of cultural attributes (such as the use of particular crop plants, domestic animals and wayfinding techniques, and of durable goods such as tools, pots and textiles) until they mixed and traded with, and presumably learned from, nearby resident Melanesians. So the picture is highly complicated. A systematic tool that could reveal hidden subgroups among similar Austronesian

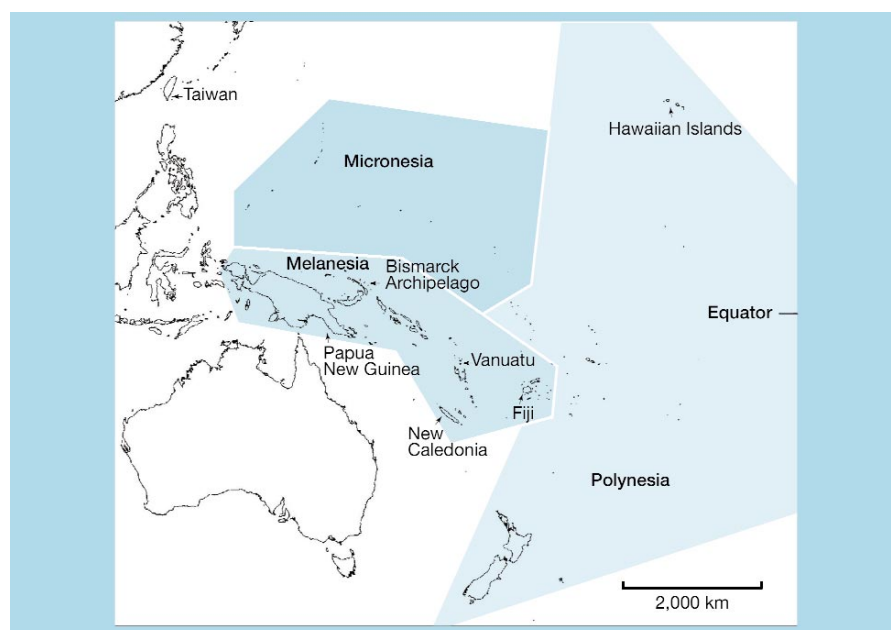


Figure 1 Peopling the Pacific. One view — the ‘express train’ theory^{5,6} based on archaeological evidence — is that, starting in Taiwan, the human colonization of the western Pacific proceeded swiftly and was complete within 2,100 years. Gray and Jordan’s analysis¹ of Austronesian languages lends support to this idea and shows how the molecular tools of evolutionary biology can be applied to linguistics. Some of the island groups (such as Marshall and Gilbert in Micronesia, and Cook and Marquesas in Polynesia) do not show on the map at this scale. Redrawn from ref. 1.

languages would be a powerful way of analysing Pacific prehistory.

This is where parsimony analysis comes in. Using a program called PAUP (ref. 3), Gray and Jordan constructed a phylogenetic tree for 77 Austronesian languages using over 5,000 lexical terms from Blust's *Austronesian Comparative Dictionary*. Their shortest tree had over 52,000 steps, and, although the tree can have several interpretations, it is a good portrayal of most of the data⁴. Gray and Jordan then set out to test a particular model for Pacific colonization, Diamond's 'express train' hypothesis^{5,6}, so-called because according to this view the 10,000-km Austronesian journey from Taiwan to its furthest reach eastwards took only about 2,100 years. This model invokes ten geographical centres or waystations for Lapita voyagers (see the map on page 1053). Gray and Jordan mapped these centres as separate character states encoding the sequence and timing of Pacific island colonization^{7,8} onto their PAUP tree using the program MacClade⁹.

The modern inhabitants of much of remote Oceania speak Austronesian languages. If the initial colonizers of the region did so, and if there have been few convergent inventions of terms, insignificant borrowing between languages and little reverse migration, then the pattern of the language tree should match the character states combining geography and earliest settlement history. So, if the actual process of settlement was strictly in accord with Diamond's 'express train', the shortest possible tree connecting the character states of distinct cultural regions to the tree of languages is predicted to be nine (ten centres minus one).

Gray and Jordan's linguistic tree was not an exact fit, but it required only 13 steps. In contrast, a random shuffling of the characters 200 times between the 77 languages required on average almost 49 steps. The close association between the character-state tree and the language tree is unlikely to be due to mere chance. Their result confirms that the pattern of language relationships among Austronesian speakers, constructed on the basis of shared lexical items, does indeed reflect one view of the history of the language speakers themselves.

Should prehistorians now encourage the wholesale adoption of parsimony analysis in comparative linguistics? It seems they have little to lose. In the 1980s, the cladistic approach of placing emphasis on similarities due to shared, derived traits, rather than the overall resemblance between organisms, stimulated evolutionary biologists to formulate testable hypotheses accounting for the origin of particular character states. The parsimony principle that it invokes fails under certain conditions — that is, it supports the incorrect tree¹⁰. Nonetheless it has proved powerful, and allows more rational debate over competing hypotheses.

There is a close connection between comparative linguistics and evolutionary biology. Both seek to account for the overall resemblance between entities that are now distinct; in both there are confounding cases of horizontal transfer of information; and both are bedevilled by spurious similarities that arise from convergence, parallelism or reversals in character states. Linguists require that comparisons of different lexicons should avoid circularity and emphasize exclusively shared innovations, and that such innovations should be appropriately weighted according to their importance. The same difficulties arise in interpreting molecules and morphologies. So given that common difficulties often have common solutions, I believe that parsimony analysis has much to offer comparative linguistics.

Much of Pacific prehistory is comparatively simple. But an accurate history of colonization and settlement cannot be reconstructed by a single parsimony analysis of one family of genes, languages or artefacts. If Diamond's express-train model is correct, a linguistic analysis restricted to the Austronesian lexical elements specific for navigation and open-ocean voyaging should yield a parsimony tree with higher consistency, more information and fewer steps than the general one presented by Gray and Jordan. Finally, another tree-building tool, the 'maximum likelihood' method, could also prove informative. Unlike a probability model with data and statistical hypotheses, likelihood deals with an outcome that has occurred (a lexicon set) and asks how likely it is that a certain tree (one based on Diamond's waystations) fits that outcome. Before it is all over, the express-train theory will surely be given the maximum-likelihood test for durability. ■

Rebecca L. Cann is in the Department of Genetics and Molecular Biology, John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii 96822, USA. e-mail: rcann@hawaii.edu

1. Gray, R. D. & Jordan, F. M. *Nature* **405**, 1052–1055 (2000).
2. Blust, R. *Symp. Ser. Inst. Linguist. Acad. Sinica* **1**, 31–94 (1999).
3. Swofford, D. L. *Phylogenetic Analysis Using Parsimony (PAUP) Version 4.0d65* (Sinauer, Sunderland, MA, 1999).
4. Huelsenbeck, J. P. *Syst. Zool.* **40**, 257–270 (1991).
5. Diamond, J. M. *Nature* **336**, 307–308 (1988).
6. Diamond, J. M. *Guns, Germs and Steel* (Cape, London, 1997).
7. Kirch, P. V. *The Lapita Peoples* (Blackwell, Cambridge, 1997).
8. Bellwood, P. *The Prehistory of the Indo-Malaysian Archipelago* 2nd edn (Univ. Hawaii Press, Honolulu, 1997).
9. Maddison, W. P. & Maddison, D. R. *MacClade: Analysis of Phylogeny and Character Variation Version 3.05* (Sinauer, Sunderland, MA, 1992).
10. Felsenstein, J. *Annu. Rev. Genet.* **22**, 521–565 (1988).

Astronomy

The cosmic origin of deuterium

Francesca Matteucci

Most of the deuterium in the Universe was thought to be created during the first three minutes after the Big Bang, along with other light elements such as hydrogen, helium and lithium. All the elements heavier than carbon were subsequently synthesized inside stars. Stars burn up primordial deuterium to form heavier elements, a process known as astration. The fact that deuterium is only ever destroyed and not created by stars makes this element a good indicator of past star formation. So, regions where there has been a high rate of stellar activity should be depleted in deuterium, unless they are resupplied with primordial gas (that is, gas with a chemical composition unchanged since the Big Bang).

On page 1025 of this issue, Lubowich *et al.*¹ report the discovery of deuterium near the centre of our Galaxy. Their measurement of deuterium (more specifically the ratio of deuterium to hydrogen, D/H) in a molecular cloud near the Galactic Centre provides crucial information about the amount of astration that has occurred over the lifetime of our Galaxy and consequently about the primordial abundance of deuterium.

The primordial abundance of deuterium is very sensitive to the density of ordinary

matter (baryons) in the early Universe and helps us to establish whether the Universe will remain open or eventually close. This is why any means of deriving the primordial deuterium abundance is important for cosmology. There are two ways of finding the primordial abundance. The first requires measurement of the abundance of deuterium in distant (high-redshift) objects, namely young protogalaxies. But these measurements only give a lower limit for the primordial abundance because the protogalaxies also contain metals (in astrophysics, metals are all the elements heavier than helium), indicating that at least some stellar processing has already taken place. The available measurements provide a rather low value² for the primordial abundance of deuterium, $(D/H)_p \approx 3.5 \times 10^{-5}$, although much higher values have been suggested³. For this reason, it is important to have an independent way of estimating the primordial abundance.

Another route is to estimate the amount of astration taking place over the lifetime of a galaxy by means of models of chemical evolution. Such models try to reconstruct the chemical history of the interstellar gas in a galaxy since the Big Bang. They do this by assuming a specific history for star for-

mation and taking into account stellar processing, in particular the processed and unprocessed material that stars release into the interstellar medium, either during their lifetime or when they die as supernovae.

Possible sources of material from outside and inside the system, such as gas flows into and out of the galaxy, are also considered. Successful models for the chemical evolution of our Galaxy suggest that the stellar disk formed by infall of mainly primordial gas, which built up faster in the inner than in the outer regions⁴. This process appears to continue today, at least in the outer regions, with evidence of gas infall from observations of incoming high-velocity clouds of neutral hydrogen⁵. The primordial abundance of deuterium is usually estimated from chemical evolution models of the local Galactic disk. Constraints on these models are the observed abundance of deuterium in the Solar System⁶ and the local interstellar medium⁷.

Most chemical evolution models of the local disk^{8,9} suggest a primordial abundance of deuterium, $(D/H)_p \approx (3.0\text{--}5.0) \times 10^{-5}$, in close agreement with the low values of D/H measured in distant (very young) objects. But these models contain many assumptions. The abundance of deuterium in the Galactic Centre is a good test for the models because, in the absence of recent infall of primordial gas, deuterium would be completely removed from the gas, unlike in the local interstellar medium.

Models devised for the centre of the Galaxy reveal that the Galactic bulge (a spherical region near the centre) formed more quickly and therefore had a greater star-formation rate than the local disk. As a result, the gas in the bulge has been enriched in heavy elements more efficiently than in the rest of the disk. The bulge and the Galactic Centre are indeed rich in metals, as demonstrated by the high iron content (almost ten times the Solar value) measured in some central stars¹⁰. A chemical evolution model¹¹ explaining such abundances also predicts a very low abundance of deuterium at the present time ($D/H \approx 3.2 \times 10^{-11}$) when a primordial abundance in the range $(3.0\text{--}5.0) \times 10^{-5}$ is assumed.

The abundance of deuterium measured by Lubowich *et al.*¹ in the Galactic Centre is the lowest ever measured in the Milky Way ($D/H \approx 1.7 \times 10^{-6}$). Nonetheless it is much larger than that predicted by models of the Galactic Centre. But the low value predicted for the centre neglects the possible infall of gas containing primordial deuterium at later times, which would increase the present D/H abundance. So the abundance measured by Lubowich *et al.* is consistent with a high rate of deuterium astration (relative to the local disk), but also requires recent infall of primordial or metal-poor material into the Galactic Centre.

When these effects are taken into account,

the deuterium abundance measured by Lubowich *et al.* is consistent with a $(D/H)_p \approx (3.0\text{--}5.0) \times 10^{-5}$, in line with previous results. For these primordial values of the deuterium abundance, the standard Big Bang model predicts that the baryonic density is only a small fraction ($\sim 3\text{--}5\%$) of the critical density necessary to close the Universe. This low baryon fraction has important cosmological implications if compared with the higher fraction ($\sim 15\text{--}25\%$) of baryons measured in galaxy clusters. On one hand it implies that most of the matter in the Universe is non-baryonic dark matter. On the other hand, it indicates that the total density of matter is only one third of the critical density needed to close the Universe. ■

Francesca Matteucci is in the Department of

Astronomy, Osservatorio Astronomico, Università Degli Studi di Trieste, via G. B. Tiepolo 11, Trieste 34131, Italy.

e-mail: matteucci@ts.astro.it

1. Lubowich, D. A. *et al.* *Nature* **405**, 1025–1027 (2000).
2. Burles, S. & Tytler, D. *Astron. J.* **114**, 1330–1336 (1997).
3. Webb, J. K. *et al.* *Nature* **388**, 250–252 (1997).
4. Chiappini, C., Matteucci, F. & Gratton, R. G. *Astrophys. J.* **477**, 765–780 (1997).
5. Wakker, B. P. *et al.* *Nature* **402**, 388–390 (1999).
6. Gloecker, G. & Geiss, J. *Nature* **381**, 210–212 (1996).
7. Linsky, J. L. *Space Sci. Rev.* **84**, 285–296 (1998).
8. Tosi, M. in *From Stars to Galaxies: The Impact of Stellar Physics on Galaxy Evolution* Vol. 98 (eds Leitherer, C. *et al.*) 299 (ASP Conf. Series, San Francisco, 1996).
9. Chiappini, C. & Matteucci, F. in *IAU Symp. 198: The Light Elements and Their Evolution* (eds da Silva, L. *et al.*) (ASP Conf. Ser., in the press).
10. McWilliam, A. & Rich, R. M. *Astrophys. J.* (suppl.) **91**, 749–791 (1994).
11. Matteucci, F., Romano, D. & Molaro, P. *Astron. Astrophys.* **341**, 458–468 (1999).

Neurobiology

Translating activity into plasticity

Aaron DiAntonio

Activity in the brain modifies the structure of individual nerves and the circuits that they form. Synapses are the main sites of communication between neurons and are a key substrate for these activity-induced changes. Long-lasting alterations in the structure and function of synapses (long-term synaptic plasticity) are crucial for the developing brain to fine-tune its wiring pattern and for the mature brain to store and process information. So identifying the molecular mechanisms by which neural activity generates long-term synaptic plasticity is central to our understanding of both normal and diseased brains.

Recently, a model for synaptic modification has emerged in which protein synthesis at synapses provides the building blocks for change^{1,2}. On page 1062 of this issue³, Sigrist and colleagues provide *in vivo* support for this model. In doing so, they establish the

fruitfly *Drosophila melanogaster* as a model organism for the study of protein synthesis and synaptic growth, creating a powerful genetic tool for understanding this mechanism of synaptic plasticity.

The first hint that proteins might be synthesized at synapses came in the mid-1980s, when it was recognized that protein-synthesizing engines such as polysomes (which translate messenger RNA into protein) accumulate at the base of tiny protrusions of postsynaptic neurons⁴. More recent experiments have shown that mRNAs encoding important synaptic proteins, such as Ca^{2+} /calmodulin-dependent kinase II (CaMKII), also localize to these 'dendrites'⁵. Biochemical experiments with cultured neurons and brain slices confirmed that dendrites can synthesize new proteins⁵. Each neuron can form synapses with many other neurons, and plasticity mechanisms regulate subsets of synapses differently.

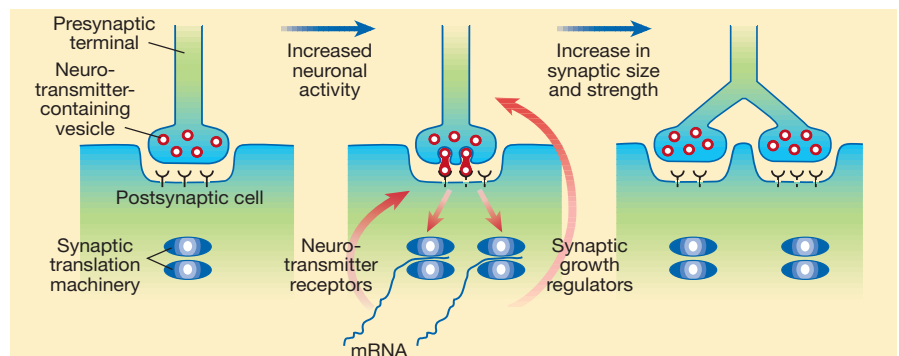


Figure 1 Local protein synthesis and the control of synaptic plasticity. Sigrist *et al.*³ show that increased neuronal activity leads to increased translation of messenger RNAs into protein in the postsynaptic cell. Synthesis of neurotransmitter receptors and synaptic growth regulators leads to an increase in the size and strength of the synapse.

So an attractive model for plasticity was that local protein synthesis at dendrites should provide a mechanism for synapse-specific changes in neuronal structure and function. If this is correct, neuronal activity should regulate synapse-specific protein synthesis, which in turn should regulate synaptic plasticity.

Several groups have found a connection between neuronal activity and the molecules that regulate the translation of mRNAs into protein. NMDA (*N*-methyl-D-aspartate) receptors are a subtype of receptor for the neurotransmitter glutamate, and regulate many forms of synaptic plasticity. Activation of these receptors in postsynaptic neurons results in the increased phosphorylation of eukaryotic elongation factor-2 (eEF-2)⁶, which is needed for protein synthesis. Synaptic activity also leads to elongation of the polyadenosine tail — a long string of adenosine nucleotides found at the end of mRNAs — of the CaMKII mRNA⁷. Both effects increase the synthesis of CaMKII protein.

Other experiments implicated local protein synthesis in the regulation of synaptic plasticity. In cultured neurons taken from *Aplysia* — a mollusc often used in neurobiology studies — protein-synthesis inhibitors delivered exclusively to the synapse can block a process called long-term facilitation⁸, which involves a long-lasting increase in synaptic strength. Similar results have been found in the mammalian hippocampus, a brain region in which local protein synthesis is required for synaptic plasticity induced by neuronal growth factors⁹.

Sigrist *et al.*³ now link together neuronal activity, postsynaptic protein translation and long-term synaptic plasticity (Fig. 1). The authors chose to study translation at the *Drosophila* neuromuscular junction. This nerve-muscle synapse controls muscle contraction, and is a favourite model system for studies of synaptic development, function and plasticity. It shares several characteristics with excitatory synapses in the vertebrate central nervous system: both are responsive to glutamate, express similar neurotransmitter receptors, have structurally similar nerve terminals, and exhibit dynamic structural and functional plasticity.

The authors looked first at two key regulators of translation, the rate-limiting initiation factor eIF4E and the polyadenosine-binding protein PABP. They found that these proteins form large aggregates within the subsynaptic reticulum, a postsynaptic specialization of the neuromuscular junction. Small amounts of these proteins are present in all cells, but the only aggregates were found at the synapse. Sigrist *et al.* analysed neuromuscular junctions by electron microscopy, and confirmed that polysomes also accumulate within the subsynaptic reticulum. They found no polysomes or translation aggregates presynaptically. In addition, mRNA encoding a postsynaptic glutamate receptor localized to the

subsynaptic reticulum. It seems that the *Drosophila* neuromuscular junction, like neuronal dendrites, is a site of protein synthesis.

Sigrist *et al.* also found that the density of the postsynaptic aggregates could be regulated by genetically manipulating their components (for example, by overexpressing PABP postsynaptically). This increase in the number of translation aggregates probably regulates the synthesis of synaptic proteins only: levels of the glutamate receptor increased under these conditions, whereas amounts of other muscle proteins, such as myosin, did not change. Interestingly, *Drosophila* mutants that show increased neuronal activity also exhibited higher levels of aggregates. These flies included those with mutations in the potassium channels *eag* and *Shaker*, or in the enzyme cyclic AMP phosphodiesterase. These data support the idea that synaptic activity can control the synthesis of synaptic proteins.

Then, to make the final link between translation and synaptic plasticity, Sigrist *et al.* looked at the consequences of genetically modifying the translation of postsynaptic proteins. Manipulations that increased the occurrence of the translation aggregates led to functionally stronger and structurally larger synapses. This result shows that postsynaptic protein synthesis is not only necessary (as shown previously^{8,9}) but is also sufficient for generating long-term plasticity. Regulated translation may be a key mediator of the effects of neuronal activity on the synapse.

These results leave us with a number of questions. For example, how does activity regulate the density of translation aggregates at the synapse? Sigrist *et al.* used indirect manipulations of activity (such as disrupting potassium channels) to investigate this problem. But does this effect normally require excitatory synaptic transmission (by glutamate and its receptors), or does it act through a different pathway? Answering this question should also provide clues as to which factors in *Drosophila* muscle regulate the protein-synthesis machinery. Most important, which proteins are regulated by translation, and how do they control the strength and size of the synapse? Now that *Drosophila* is established as a model system for the study of synaptic protein synthesis, one hopes that the oft-mentioned power of genetic analysis will translate into rapid progress towards understanding synaptic plasticity. ■

Aaron DiAntonio is in the Department of Molecular Biology and Pharmacology, Washington University School of Medicine, Campus Box 8103, 660 South Euclid Avenue, St Louis, Missouri 63110, USA.

e-mail: diananton@pcg.wustl.edu

1. Wells, D. G. *et al.* *Curr. Opin. Neurobiol.* **10**, 132–137 (2000).
2. Schuman, E. M. *Neuron* **23**, 645–648 (1999).
3. Sigrist, S. J. *et al.* *Nature* **405**, 1062–1065 (2000).
4. Steward, O. & Falk, P. M. *J. Neurosci.* **6**, 412–423 (1986).
5. Steward, O. *Neuron* **18**, 9–12 (1997).
6. Scheetz, A. J. *et al.* *Nature Neurosci.* **3**, 211–216 (2000).
7. Wu, L. *et al.* *Neuron* **21**, 1129–1139 (1998).
8. Martin, K. C. *et al.* *Cell* **91**, 927–938 (1997).
9. Kang, H. & Schuman, E. M. *Science* **273**, 1402–1406 (1996).

Daedalus

Tradesman's entrance

As a retailing device, the Internet is wildly oversold. Compared to old-style mail order, it merely offers one delivery delay instead of two. Against that, anything you order from anybody is instantly noted by any number of interested parties, who will use the information to pester you for the rest of time. And, of course, when the courier turns up with the long-awaited parcel, you will be out.

In this connection, Daedalus remarks what a brilliant invention the front-door letter-box is. It lets a postman deliver mail even if you are not at home. And it is wonderfully secure. A thief cannot get in through it, nor remove a letter that has been pushed in, or even tell if one has been delivered. For e-commerce to take off, says Daedalus, we need an equivalent domestic 'parcel-box'. It is a low-tech problem beside the digital wonders of the Internet. But it does need to be solved.

The first requirement is a standard parcel, acceptable by all parcel-boxes. The standard freight container, 10 feet by 10 feet by 20 feet, is rather big for the job. Daedalus recommends something the size of a small suitcase. Like a freight container, it would be freely re-useable, and would accept most consumer items. Even a large grocery order (say) could be fitted into a number of such 'minicon' containers.

The matching minicon-receiver will need some subtlety. It must have a door which opens only to accept a minicon, and then swallows it irretrievably. Inside, a conveyor system must let it pack and stack many minicons in succession. It should be able to read a coded delivery note, imprint it with a valid acceptance signature, and log the reception for the householder. It must resist entry by a grappling device, or even a diminutive thief, and should not be stealable itself. And it must be easily 'retrofitted' to buildings of all kinds.

DREADCO engineers are devising a sandpit which, when activated by a special deliverer's key, is fluidized by a stream of air. A minicon laid on it would then sink in and vanish from sight. At other times nothing could be made to pass in either direction. But the field is still wide open. A rival company may leap in to establish the standard minicon and its receiver. Whoever does it will catalyse a commercial revolution. Mail, groceries, consumer durables and trivials, all will flow as freely and safely as information. And many bulky consumer items will be usefully redesigned and miniaturized to fit into a minicon.

David Jones

Conditioning and opiate withdrawal

The amygdala links neutral stimuli with the agony of overcoming drug addiction.

The habitual behaviour of many opiate addicts means that the aversive symptoms of drug withdrawal are frequently experienced in specific environments¹. Stimuli in these environments acquire some of the negative affective properties of opiate withdrawal through pavlovian conditioning, and subsequent exposure to these stimuli alone can then elicit a conditioned withdrawal response that is often associated with drug craving, drug seeking, and relapse back to compulsive drug use after a long abstinence^{1,2}. We show here that selective excitotoxic lesions of the basolateral amygdala, part of the limbic forebrain, prevent the development of conditioned withdrawal in morphine-dependent rats.

Addiction to opiate drugs, such as heroin, depends not only upon their positive reinforcing and hedonic effects, but also upon escape from, or avoidance of, the negative, aversive consequences of withdrawal³. The nucleus accumbens, prefrontal cortex and central nucleus of the amygdala, together with their associated connections with the midbrain, are all involved in the acute, positive reinforcing or rewarding effects of opiates and other drugs of abuse^{3,4}. It has been proposed that neuroadaptations within these brain regions during the development of drug dependence contribute to the negative affective components of withdrawal that motivate continued compulsive drug use³⁻⁵.

By contrast, it is less clear how stimuli in the drug-taking environment become associated with both the positive and negative effects of abused drugs and thereby impact on the persistence of drug-seeking behaviour, although the basolateral amygdala has been implicated in the formation of such conditioned associations^{6,7}.

We assessed the aversive properties of a conditioned stimulus paired with withdrawal, precipitated by the opiate-receptor antagonist naloxone, in morphine-dependent rats. We tested the ability of the conditioned stimulus alone to disrupt an appetitive behavioural response, namely responding for food by hungry rats⁸. Male Wistar rats were trained to press a lever for a food reward under a fixed-ratio schedule and then received either selective, quinolinic acid-induced lesions of the basolateral amygdala or sham lesions (Fig. 1a). After responding for food was re-established, two 75-mg morphine pellets were implanted subcutaneously (s.c.) to induce dependence. Five days later, both sham and lesioned rats received four pairings of naloxone (0.03 mg kg⁻¹, s.c.)-precipitated withdrawal with a tone/light

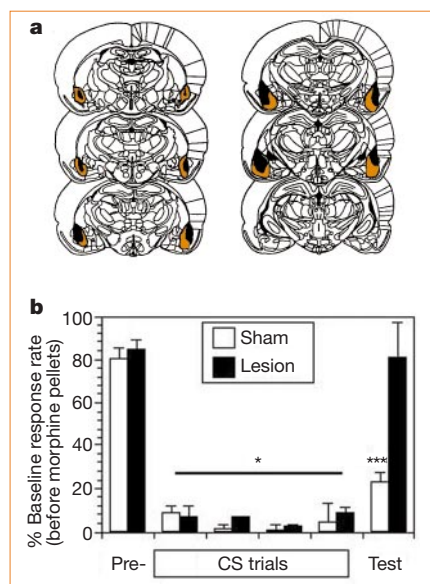


Figure 1 Basolateral amygdala lesions significantly attenuate the conditioned withdrawal produced by a tone/light conditioned stimulus (CS) paired with naloxone during four conditioning trials. **a**, Lesions of the basolateral amygdala were made in rats anaesthetized with halothane (1–4%) by infusing 0.09 M quinolinic acid (two injections in each hemisphere, 0.3 μ l each at coordinates: anterior–posterior, –2.3 and –3.0 mm from bregma; lateral: $\pm$ 4.6 mm; vertical, –7.3 mm from the dural surface; incisor bar, –3.3)⁶. Sham lesions were made by infusing 0.1 M phosphate buffer (pH 7.0–7.3) vehicle only. In this representation of basolateral amygdala lesions, shaded areas represent the smallest (black) and largest (orange) extent of neuronal loss mapped onto coronal sections of the rat brain at anterior–posterior levels –1.8 mm (top left) to –4.16 mm (bottom right) relative to bregma. Neuronal loss was significant in the parvocellular and magnocellular subdivisions of the basal amygdaloid nucleus, in the accessory basal nucleus and in the lateral nucleus of the amygdala. Further experimental details are available from the authors. **b**, Data represent mean ($\pm$ s.e.m.) per cent of baseline response rates over the entire 20-min test session for the ‘lesion’ ($n=7$) and ‘sham’ ($n=6$) groups. The sham and lesion groups show comparable suppression of response rates during conditioning trials when naloxone is present (CS trials, $*P<0.05$ vs pre-conditioning (‘pre-’, left) baseline). Significant suppression of responding by the CS alone at ‘test’ is seen in the sham but not the lesion group ($*P<0.05$ vs pre-conditioning baseline; $**P<0.05$ lesion vs sham group at test). The dose of naloxone used (0.03 mg kg⁻¹) produces almost complete suppression of operant responding, but is well below the half-maximal effective dose for most somatic signs of opiate withdrawal¹².

compound stimulus (7 kHz, 85 dB tone plus house light on for 5 s, then off for 2 s) during consecutive daily 20-min sessions.

The baseline response rates of sham (98.2 ± 7.5) and lesioned (91.4 ± 9.4) rats after morphine pellet implantation did not differ; both groups showed a profound, unconditioned withdrawal response (suppression of response rates within 5 min of

the naloxone injection; Fig. 1b). However, during a test session 24 hours after the final conditioning trial, the robust conditioned suppression of response rates seen in the sham-operated rats following presentation of the withdrawal-associated conditioned stimulus alone was completely abolished in rats with basolateral amygdala lesions (Fig. 1b). These results show that the basolateral amygdala is not required for the primary aversive effects of opiate withdrawal, but that it is critical for the association of the negative affective state of withdrawal with previously neutral environmental stimuli.

Our results parallel those from studies of conditioning to positive affective states. Although lesions of the basolateral amygdala do not affect the primary reinforcing effects of natural reinforcers such as sucrose, water or sexual interaction, or drugs such as heroin and cocaine^{6,9}, they markedly impair the process by which environmental stimuli paired with such rewards gain control over instrumental behaviour, by acting as conditioned reinforcers^{7,9}. These observations suggest that the amygdala mediates specific forms of learning, whereby environmental events are associated with both positive and negative primary reinforcers, thus mediating adaptive behavioural responses during future encounters with such predictive cues^{7,9}.

The cellular mechanisms within the basolateral amygdala that mediate conditioned withdrawal processes and their interaction with other structures of the brain's reinforcement circuitry still need to be identified. Probable participants include the nucleus accumbens, central amygdala and bed nucleus of the stria terminalis, some of which are components of the ‘extended amygdala’. These are widely accepted as primary sites mediating the positive and negative reinforcing effects of drugs of abuse^{4-7,10}, and they are also a major target for projections from the basolateral amygdala¹¹. We have identified the basolateral amygdala as a key element within this circuitry that may sustain compulsive opiate drug use and contribute to drug craving and to relapse by mediating the motivational impact of conditioned withdrawal.

G. Schulteis*†, S. H. Ahmed†, A. C. Morse*†, G. F. Koob†, B. J. Everitt†

*Department of Anesthesiology, VAMC 125, UC San Diego School of Medicine and San Diego VA Medical Center, 3350 La Jolla Village Drive, San Diego, California 92161-5085, USA

†Department of Neuropharmacology, CVN-7, Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

‡Department of Experimental Psychology,
University of Cambridge, Downing Street,
Cambridge CB2 3EB, UK
e-mail: bje10@cus.cam.ac.uk

1. Wikler, A. *Arch. Gen. Psychiatry* **28**, 611–616 (1973).
2. O'Brien, C. P., Testa, T., O'Brien, T. J., Brady, J. P. & Wells, B. *Science* **195**, 1000–1002 (1977).
3. Koob, G. F. & Le Moal, M. *Science* **278**, 52–58 (1997).
4. Schulteis, G. & Koob, G. F. *Neurochem. Res.* **21**, 1437–1454 (1996).
5. Koob, G. F., Wall, T. L. & Bloom, F. E. *Psychopharmacology* **98**, 530–534 (1989).
6. Whitelaw, R. B., Markou, A., Robbins, T. W. & Everitt, B. J.

- Psychopharmacology* **127**, 213–224 (1996).
7. Everitt, B. J. et al. *Ann. NY Acad. Sci.* **877**, 412–438 (1999).
8. Baldwin, H. R. & Koob, G. F. *Neuropsychopharmacology* **8**, 15–21 (1993).
9. Everitt, B. J., Cardinal, R. N., Hall, J., Parkinson, J. & Robbins, T. W. in *The Amygdala: A Functional Analysis* (ed. Aggleton, J.) (Wiley, New York, in the press).
10. Aston-Jones, G., Delfs, J. M., Druhan, J. & Zhu, Y. *Ann. NY Acad. Sci.* **877**, 486–498 (1999).
11. Groenewegen, H. J., Wright, C. I. & Beijer, A. V. J. *Progr. Brain Res.* **107**, 485–511 (1996).
12. Schulteis, G., Markou, A., Gold, L. H., Stinus, L. & Koob, G. F. *J. Pharmacol. Exp. Ther.* **271**, 1391–1398 (1994).

Biotechnology

Enzymatic production of biohydrogen

Although in theory the amount of hydrogen that could be generated from renewable sources of energy such as cellulose (a polymer of glucose) is vast¹, only 16–24% of the maximum stoichiometric yield of hydrogen from glucose (about 12 mol H₂ per mol glucose) is typically achieved by biological methods². Here we show that the enzymes of the oxidative pentose phosphate cycle^{3–5} can be coupled to hydrogenase purified from the bacterium *Pyrococcus furiosus*, one of only a few hydrogenases that use NADP⁺ as the electron carrier⁶, to generate 11.6 mol H₂ per mol glucose-6-phosphate. Hydrogen produced by this pathway is the major product, unlike that produced by intermediate metabolic pathways of bacterial fermentation, and therefore has important practical implications for biohydrogen production⁷.

The oxidative branch of the pentose phosphate cycle containing glucose-6-phosphate (G6P) dehydrogenase and 6-phosphogluconate (6PG) dehydrogenase, converts G6P to ribulose-5-phosphate (Ru5P), with the formation of 2 mol NADPH and 1 mol CO₂. Incubation of G6P and NADP⁺ with these dehydrogenases and the *P. furiosus* hydrogenase generates the maximum possible stoichiometric yield of H₂, 2 mol per mol G6P (Fig. 1, lower curve). CO₂ production⁸ was estimated to be 0.85 μmol, agreeing quite well with the theoretically predicted yield (1 μmol). The rate of H₂ evolution reached a maximum of about 300 nmol h^{−1} and then declined to zero after 30 h.

However, when all the enzymes of the oxidative branch of the pentose phosphate cycle apart from 6-phosphogluconolactonase were mixed with the bacterial hydrogenase, NADP⁺ and G6P, 11.6 mol H₂ per mol G6P were generated. In this case, the rate of H₂ evolution reached a maximum of 425 nmol h^{−1} and declined to zero after 80 h (Fig. 1, top curve). This is an important result because the generation of 11.6 mol H₂ from 1 mol G6P represents 97% of the maximum stoichiometric yield

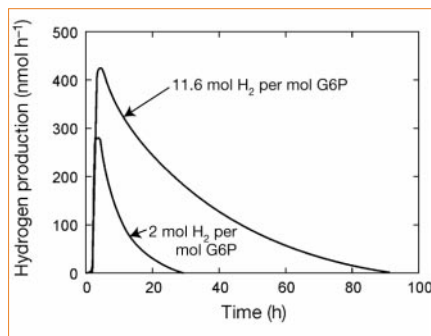


Figure 1 Hydrogen production by enzymes of the oxidative pentose phosphate cycle. Lower curve, the enzymatic conversion of glucose-6-phosphate (G6P) to ribulose-5-phosphate (Ru5P) with concomitant evolution of H₂. The reaction mixture contains 2 μmol G6P, 1 unit each of G6P and 6-phosphogluconic (6PG) dehydrogenases, 63 units of *Pyrococcus furiosus* hydrogenase (University of Georgia, Athens) and 4 μmol NADP⁺ in 2.0 ml 50 mM sodium phosphate buffer, pH 7.5, at 40 °C. The molar yield of H₂ per mol G6P was 2. Top curve, the enzymatic conversion of G6P to H₂ by enzymes of the oxidative and regenerative branch of the pentose phosphate cycle. The reaction mixture contains 1 μmol G6P, 1 unit each of the enzymes of the oxidative pentose phosphate cycle, 1 μmol thiamine pyrophosphatase (co-carboxylase), 1.0 μmol NADP⁺, and 63 units of *P. furiosus* hydrogenase in 2.0 ml 0.2 M HEPES buffer, pH 7.5, containing 10 mM MgCl₂ at 30 °C. The molar yield of H₂ per mol G6P was 11.6. The starting pH was 7.4; after 92 h of reaction, the pH was 7.2. Hydrogen evolution was measured using an in-line hydrogen sensor with a flow-detection system, as described⁹. Mesophilic forms of the pentose phosphate enzymes were purchased from Sigma.

of H₂ from G6P attainable by this pathway. To our knowledge, this is the first time such a high yield of H₂ from glucose has been obtained by any biological process.

This remarkable yield of hydrogen from the oxidative pentose cycle means that the reaction is proceeding almost to completion as a result of the hydrogenase oxidizing NADPH as soon as it is formed and in effect removing it from the reaction; the hydrogen evolved is then swept out of the reaction vessel by a helium carrier gas⁹. The ratio of reductant to CO₂ generated by the complete pentose phosphate cycle was always about 2:1 during the course of the reaction. The kinetics of the individual reactions in this complex system remain to be determined.

We detected no G6P in the reaction mixture at the end of the reaction, as would be expected if the starting G6P was completely

oxidized by G6P dehydrogenase. This would not hold for other intermediates of the cycle, including ribose-5-phosphate, xylulose-5-phosphate and glyceraldehyde-3-phosphate, because the number of molecules of these intermediates would eventually become too low during the course of the reaction to complete the cycle. In other words, it is not possible to obtain as much as 12 mol H₂ from 1 mol glucose or G6P, although a yield approaching 12 is feasible.

The effect of 6-phosphogluconolactonase, which catalyses the hydrolysis of the lactone ring to free 6PG, on the kinetics of hydrogen production was not determined. However, as 6-phosphogluconolactone spontaneously hydrolyses to 6-phosphogluconic acid in aqueous solution, resulting in a stable equilibrium, the presence of 6-phosphogluconic dehydrogenase would shift the position of equilibrium in favour of 6-phosphogluconate formation.

Calculations using published thermodynamic data^{10,11} revealed a change in standard Gibbs free energy ($\Delta_r G^\circ$) of −31.97 kJ mol^{−1} for the complete conversion of glucose to 12 mol H₂. Under standard conditions, consumption of 12 mol H₂, for example by a fuel cell, can generate a maximum of 2,845.13 kJ of free energy to power a process: this recoverable energy is what makes the process energetically efficient. So even if we take into account the energy required to make G6P (−19 kJ mol^{−1}), 98% of the potential energy of hydrogen combustion is recoverable.

In any biotechnological process involving the oxidative pentose phosphate cycle, glucose derived from cellulose must first be phosphorylated to G6P, a reaction that involves the ATP-dependent enzyme hexokinase; the ATP cofactor is available in bulk commercially. Enzymatic production of G6P is feasible on an industrial scale, with ATP being regenerated from ADP and inorganic phosphate so that the reaction can continue by using the thermostable enzymes acetate kinase and adenylate kinase in immobilized form^{12,13}.

If thermally stable forms of the enzymes of the pentose phosphate cycle could be found that would function at a higher reaction temperature, then presumably the reaction could be speeded up. The genes encoding several of the enzymes of the oxidative pentose phosphate cycle from the archaeon *Methanococcus jannaschii* are commercially available from the American type cell culture collection. Such stable forms of these enzymes would be compatible with our hydrogenase from the hyperthermophile *P. furiosus*, which works at an optimum temperature of 85 °C and is much less active at 40 °C. No mesophilic form of this hydrogenase is available that uses NADPH as an electron donor.

Although the oxidative pentose phos-

phate cycle generates 6 mol CO₂ per mol glucose, it can still be considered as 'green' technology and not a potential contributor to global warming, because it recycles all CO₂ reduced to sugar by the process of photosynthesis — rather than using fossil fuels to produce fuels and chemicals, which increase CO₂ in the atmosphere.

Jonathan Woodward, Mark Orr, Kimberley Cordray, Elias Greenbaum
Chemical Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6194, USA
e-mail:woodwardj@ornl.gov

- Woodward, J. *et al.* *Nature Biotechnol.* **14**, 872–874 (1996).
- Keasling, J. D. *et al.* in *Biohydrogen* (ed. Zaborsky, O. R.) 87–97 (Plenum, New York, 1998).
- Horecker, B. L. & Smyrniotis, P. Z. *Arch. Biochem.* **29**, 232–241 (1950).
- Couri, D. & Racker, E. *Arch. Biochem. Biophys.* **83**, 195–205 (1959).
- Lehninger, A. L. *Biochemistry* 2nd edn (Worth, New York, 1975).
- Ma, K., Zhou, Z. H. & Adams, M. W. W. *FEMS Microbiol. Lett.* **122**, 245–250 (1994).
- Benemann, J. R. *Nature Biotechnol.* **14**, 1101–1103 (1996).
- Lee, J. W. *et al.* *Science* **273**, 364–367 (1996).
- Greenbaum, E. *Photobiophys. Photobiophys.* **8**, 323–332 (1984).
- Stull, D. R., Westrum, E. F. & Sinke, G. C. in *The Chemical Thermodynamics of Organic Compounds* 680 (Wiley, New York, 1969).
- Lide, D. R. & Frederikse, H. P. R. *CRC Handbook of Chemistry and Physics* 5-16, 5-24 (CRC, London, 1995).
- Chenault, H. K. & Whitesides, G. M. *Appl. Biochem. Biotechnol.* **14**, 147–197 (1987).
- Nakajima, H. *et al.* *ACS Symp. Ser.* **466** (eds Bednarski, M. D. & Simon, E. S.) 111–120 (Am. Chem. Soc., Washington DC, 1991).

Population genetics

Malaria susceptibility and CD36 mutation

A critical step in infection by *Plasmodium falciparum*, the microorganism that causes the most severe form of malaria, is the adhesion of parasitized red blood cells to capillary endothelium. The human protein CD36 is a major receptor for *P. falciparum*-infected red blood cells^{1,2} and may contribute to the disease by sequestering infected red blood cells^{1,2} and inhibiting the immune response to the parasite³. We have found that African populations contain an exceptionally high frequency of mutations in CD36. Unexpectedly, these mutations that cause CD36 deficiency are associated with susceptibility to severe malaria, suggesting that the presence of distinct CD36 mutations in Africans and Asians^{4–6} is due to some selection pressure other than malaria.

Many people of African origin, but few whites, are deficient in CD36 (ref. 6). Sequence analysis in five CD36-deficient African Americans showed that all subjects were homozygotes or compound heterozygotes for nonsense or frameshift mutations (Table 1). These mutations encode short CD36 proteins predicted to lack the carboxy-terminal transmembrane domain⁷.

Table 1 Molecular basis of CD36 deficiency in African Americans

Patient	Age (yr)	Clinical presentation	Type	Mutations
D.D.	15	Peripheral neuroepithelioma	I	T1264G/T1264G
R.G.	21	Thrombocytopenic offspring	I	T1264G/T1264G
C.S.	41	Acute myeloid leukaemia	I	T1264G/T1264G
M.A.	44	Healthy blood donor	II	T1264G/1144delG
S.B.	38	Healthy blood donor	ND	T1264G/1530ins14bp

CD36 mutations were identified by direct sequencing of promoters and exons amplified by the polymerase chain reaction (PCR) from genomic DNA. All CD36-deficient patients had the NaK⁺-negative platelet phenotype. Type I patients had no CD36 detectable on monocytes⁸. Cloned PCR products showed that T1264G and 1144delG are non-allelic in M.A. (data not shown). Nucleotide numbering is according to Arnesilla⁷. ND, not determined. Primer sequences are available from the authors.

The T1264G stop mutation in exon 10, which accounts for 80% of mutant alleles detected, was present in all five subjects.

Additional sequencing in 18 patients with severe malaria and 18 controls from the Gambia revealed two further mutations that encode similarly truncated CD36 proteins: G1439C+1444delA, an Ala→Pro substitution and frameshift in exon 12, was detected in two malaria patients, and the point deletion 990delG was found in a single patient.

We measured the frequencies of T1264G and G1439C+1444delA — the only mutations with allele frequencies of more than 1% in Gambians — in three African populations. In the combined Gambian and Kenyan data sets, the frequency of all detected mutations was higher in patients with severe malaria than in controls ($P=0.01$), owing to an excess of mutant genotypes in patients with cerebral malaria (Table 2). In Kenyans, the homozygous state for the T1264G stop mutation was also associated with susceptibility to cerebral malaria ($P=0.04$). We also found a high frequency of these mutations in a population-based cohort of healthy UK Afro-Caribbeans⁸ (T1264G allele frequency, 9.9%; G1439C+1444delA, 0.3%).

The homozygous state for T1264G (the commonest mutation) is associated with complete CD36 deficiency on platelets and monocytes⁴ (Table 1). This mutation may have arisen on a single founder chromo-

some because it is in strong linkage disequilibrium with a single allele of the CD36 intron-3 microsatellite⁹ in all four African populations tested (data not shown). CD36 mutations in Africans appear to be non-randomly distributed: the 314 alleles that we sequenced partly or entirely contained seven nonsense or missense mutations (the two missense mutations are not shown) but no silent substitutions, suggesting selection for non-synonymous mutations. All seven mutations lie between the malaria-binding and C-terminal transmembrane domains^{2,7}.

The exceptionally high frequency (7.5–18.5%) of the heterozygous or homozygous state for these CD36 truncation mutations is significantly associated with susceptibility to severe malaria (particularly cerebral malaria; Table 2), perhaps because of reduced parasite sequestration in organs outside the brain.

Although CD36 is known to affect several pathological steps in malaria infection, our results suggest that the outcome of infection is not determined predominantly by CD36-mediated suppression of the immune response³. CD36 mutations may also influence susceptibility to mild malaria cases, and parasites may show temporal or geographical heterogeneity in CD36 binding and pathogenicity, but these issues are not investigated here.

Our results indicate that CD36 mutations in Africans may be maintained by unidenti-

Table 2 Frequencies of CD36 mutations in Gambians and Kenyans

Genotype	All malaria	Cerebral	Anaemia	Controls
Gambians	(n=415)	(n=291)	(n=157)	(n=430)
Exon 10				
WT	95.0	94.5	96.8	97.2
T1264G/WT	5.0	5.5	3.2	2.6
T1264G/T1264G	0	0	0	0.2
Exon 12				
WT	95.9	96.0	94.9	96.3
G1439C+1444delA/WT	4.1	4.0	5.1	3.7
Total mutant allele frequency	4.5*	4.6†	4.1	3.3
Kenyans	(n=183)	(n=97)	(n=86)	(n=331)
Exon 10				
WT	79.2	76.3	82.6	82.8
T1264G/WT	19.1	20.6	17.4	16.9
T1264G/T1264G	1.6	3.1‡	0	0.3
Total mutant allele frequency	11.2*	13.4†	8.7	8.8

CD36 genotype frequencies (%) in patients with severe malaria (as defined previously in refs 10,11) and controls from the Gambia¹⁰ and Kenya¹¹. The G1439C+1444delA mutation was not detected in Kenyans. Kenyan cases had either cerebral malaria or severe anaemia, whereas 33 Gambian subjects had both cerebral malaria and severe anaemia. T1264G genotypes were generated by PCR using primers NdelF 5'-TTGTTCTTCATCCATATGGCCTTTGCCTCTCCAG-3', which contains an engineered *NdeI* site (bold), and NdelR 5'-CTGTGCTACTGAGGTATTATAC-3'. G1439C+1444delA genotypes were generated by amplification-refractory mutation assay using wild-type forward primer 5'-ATAACTGGATTCACCTTACAAATTTCGAAGA-3', and mutant forward primer 5'-ATAACTGGATTCACCTTACAAATTTCGAAGC-3'. The common reverse primer was 5'-TACTTCATGTTCTTACAAAGAGATACAG-3'. Mutant G1439C+1444delA alleles were confirmed by direct sequencing. WT, wild-type.

*Mantel-Haenszel (M-H) odds ratio = 1.53 for all malaria patients vs controls in combined Gambian and Kenyan data sets; 95% CI, 1.09–2.15, $P=0.01$. †M-H odds ratio = 1.49 for cerebral patients vs controls in combined data sets; 95% CI, 1.01–2.20, $P=0.04$. The results were robust to M-H stratification for ethnic group.

‡ $\chi^2=6.3$, Exact $P=0.04$ for homozygote frequency in cerebral patients vs controls in Kenyans.

fied selection pressures, probably a prevalent infection other than malaria. The high frequency of CD36 deficiency in other races^{4,5} may have a similar explanation.

Timothy J. Aitman*, Lisa D. Cooper*, Penny J. Norsworthy*, Faisal N. Wahid*, Jennefer K. Gray*, Brian R. Curtis†, Paul M. McKeigue‡, Dominic Kwiatkowski§, Brian M. Greenwood§, Robert W. Snow||, Adrian V. Hill¶, James Scott*

*Molecular Medicine Group, MRC Clinical Sciences Centre, and Imperial College Genetics and Genomics Research Institute, Hammersmith Hospital, London W12 0NN, UK
e-mail: t.aitman@csc.mrc.ac.uk

†Blood Center of Southeastern Wisconsin, Medical College of Wisconsin, Milwaukee, Wisconsin 53201-2178, USA

‡Department of Epidemiology and Population Sciences, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

§MRC Laboratories, Fajara, The Gambia

||KEMRI Coastal Research Unit, Kilifi, Kenya

¶Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK

1. Ockenhouse, C. F., Tandon, N. N., Magowan, C., Jamieson, G. A. & Chulay, J. D. *Science* **243**, 1469–1471 (1989).
2. Baruch, D. I., Ma, S. C., Pasloske, B., Howard, R. J. & Miller, L. H. *Blood* **94**, 2121–2127 (1999).
3. Urban, B. C. et al. *Nature* **400**, 73–77 (1999).
4. Kashiwagi, H. et al. *J. Clin. Invest.* **95**, 1040–1046 (1995).
5. Urwijitaroon, Y., Barusru, S., Romphruk, A. & Puapairoj, C. *Transfusion* **35**, 868–870 (1995).
6. Curtis, B. R. & Aster, R. H. *Transfusion* **36**, 331–334 (1996).
7. Armesilla, A. L. & Vega, M. A. *J. Biol. Chem.* **269**, 18985–18991 (1994).
8. McKeigue, P. M., Shah, B. & Marmot, M. G. *Lancet* **337**, 382–386 (1991).
9. Lipsky, R. H., Ikeda, H. & Medved, E. S. *Hum. Mol. Genet.* **3**, 217 (1994).
10. Hill, A. V. S. et al. *Nature* **352**, 595–600 (1991).
11. Snow, R. W. et al. *Trans. R. Soc. Trop. Med. Hyg.* **87**, 386–390 (1993).

Taxonomy

Species status of hybridizing oaks

The two widespread species of oak tree in Europe, *Quercus robur* L. and *Q. petraea* (Matt.) Liebl., hybridize extensively, calling their taxonomic status into question. Here we use microsatellite DNA, a highly informative genetic marker, to show that *Q. robur* and *Q. petraea* are discrete taxonomic units despite this intensive hybridization. Furthermore, individual oaks can be assigned to separate species.

Oaks are a classic example of a taxonomic group that has significantly challenged existing species concepts¹, of which the 'biological species concept' is the most popular. Two European species, *Q. robur* and *Q. petraea*, hybridize without any significant mating barriers². Despite the high gene flow between them, the two species show clear differences in leaf and fruiting structures³. The two species are largely sympatric, but there are

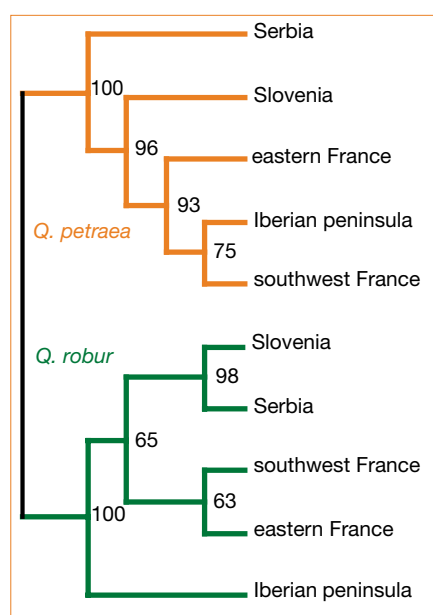


Figure 1 Dendrogram of 10 European *Quercus robur* and *Q. petraea* populations (162 individuals) based on the proportion of shared alleles at 20 microsatellite loci. One population from each species was sampled at five locations. The dendrogram was computed using a UPGMA approach in PHYLIP¹². Numbers are bootstrap support values.

some habitat differences: *Q. robur* grows in wetter and more alkaline habitats, whereas *Q. petraea* is more drought resistant.

In contrast to morphological characters and ecological preferences, it is difficult to find any interspecific differences using molecular markers. Chloroplast markers were found to be polymorphic, but could not discriminate between *Q. robur* and *Q. petraea*^{4,5}. Similarly, the two species show very similar RAPD and allozyme frequencies^{6–8}; nor did ITS sequences, a popular marker for species determination, reveal any interspecific differences (G.M., C.C.F. and C.S., manuscript submitted).

The taxonomic status of oaks has implications for forest management. Given the two species' different habitat preferences, the European Community, seeking to avoid the use of maladapted seed, has established guidelines for the trade of oak seed material (71/161/EEG, 30.3.1971). According to these, each seed lot must not contain more than 0.1% of seeds from another species. But without unambiguous criteria for taxonomic classification, this guideline cannot be put into practice.

We used microsatellite analysis, which has been shown to discriminate between closely related species⁹, to address the question of species status in *Q. robur* and *Q. petraea*. To account for geographic differences, we sampled one population (approximately 16 individuals) from each species at five locations that covered almost all the species' European range. If *Q. robur* and *Q. petraea* are good taxonomic units, then populations should cluster according to

species rather than geographic origin. Figure 1 shows a dendrogram based on the proportion of shared alleles¹⁰ at 20 microsatellite loci. In accordance with the assumption that *Q. robur* and *Q. petraea* are separate taxonomic units, all populations of the same species group together. This separation between the two species has very strong bootstrap support (100%). *F*-statistics also indicate a significant difference between the two species ($P < 0.01$), although there is also a significant component of variation due to the geographic origin of the population ($P < 0.01$).

In the light of the EC's requirement for unmixed seed lots, we tested whether the genotypic information from the 20 loci typed for both species is sufficient to identify the species of an individual. Eighty-one oak samples from Ireland, which were not included in the previous analysis, were typed using the same set of microsatellites. Using an assignment test based on allele frequencies¹¹, we determined the species status of those 46 samples that had either a clear *Q. robur* or *Q. petraea* phenotype, based on three characters of leaf morphology (petiole length/lamina length, auricle development and mid-rib pubescence). For no individual was the incorrect species assigned. Morphological and molecular evidence agreed in 78% of cases. In 22% of the cases, the molecular analysis was either not informative (7%) or indicated a hybrid status (15%).

Our molecular analysis demonstrates that *Q. robur* and *Q. petraea* are separate taxonomic units, which can be designated by the use of microsatellites. The important question remains as to how the species differences are maintained despite the high levels of interspecific gene flow.

Graham Muir*†, Colin C. Fleming*, Christian Schlotterer†

*Department of Applied Plant Science, Queen's University of Belfast, Newforge Lane, Belfast BT9 5PX, UK

†Institut für Tierzucht und Genetik, Veterinärmedizinische Universität Wien, Josef-Baumgarten Gasse 1, 1210 Wien, Austria
e-mail: Christian.Schlotterer@vu-wien.ac.at

1. Burger, W. C. *Taxon* **24**, 45–50 (1975).
2. Bacilieri, R., Ducousso, A., Petit, R. J. & Kremer, A. *Evolution* **50**, 900–908 (1996).
3. Rushton, B. S. *Irish Forestry* **40**, 52–77 (1983).
4. Ferris, C., Oliver, R. P., Davy, A. J. & Hewitt, G. M. *Mol. Ecol.* **2**, 337–344 (1993).
5. Dumolin-Lapègue, S., Kremer, A. & Petit, R. J. *Evolution* **53**, 1406–1413 (1999).
6. Bodénès, C., Joandet, S., Laigret, F. & Kremer, A. *Heredity* **78**, 433–444 (1997).
7. Samuel, R. *Plant Syst. Evol.* **217**, 137–146 (1999).
8. Zanetto, A., Roussel, G. & Kremer, A. *Forest Genet.* **1**, 111–123 (1994).
9. Harr, B., Weiss, S., David, J. R., Brem, G. & Schlotterer, C. *Curr. Biol.* **8**, 1183–1186 (1998).
10. Minch, E., Ruiz-Linares, A., Goldstein, D., Feldman, M. & Cavalli-Sforza, L. L. <http://lotka.stanford.edu/microsat/microsat.html> (1995).
11. Cornuet, J.-M., Piry, S., Luikart, G., Estoup, A. & Solignac, M. *Genetics* **153**, 1989–2000 (1999).
12. Felsenstein, J. *Cladistics* **5**, 164–166 (1989).

Effect of aquaculture on world fish supplies

Rosamond L. Naylor*, Rebecca J. Goldberg†, Jurgenne H. Primavera‡, Nils Kautsky§||, Malcolm C. M. Beveridge¶, Jason Clay#, Carl Folke§||, Jane Lubchenco*, Harold Mooney* & Max Troell§||

* Stanford University, Institute for International Studies, Encina Hall 400E, Stanford University, Stanford, California 94305-6055, USA

† Environmental Defense, 257 Park Avenue South, New York, New York 10010, USA

‡ Aquaculture Department, Southeast Asian Fisheries Development Center, Tigbauan, Iloilo, 5021, Philippines

§ Department of Systems Ecology, Stockholm University, S-106 91 Stockholm, Sweden

|| The Beijer Institute, Stockholm, Sweden

¶ Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, UK

World Wildlife Fund, 1250 24th Street NW, Washington DC 20037, USA

* Department of Zoology, Oregon State University, Corvallis, Oregon 97331-2914, USA

Global production of farmed fish and shellfish has more than doubled in the past 15 years. Many people believe that such growth relieves pressure on ocean fisheries, but the opposite is true for some types of aquaculture. Farming carnivorous species requires large inputs of wild fish for feed. Some aquaculture systems also reduce wild fish supplies through habitat modification, wild seedstock collection and other ecological impacts. On balance, global aquaculture production still adds to world fish supplies; however, if the growing aquaculture industry is to sustain its contribution to world fish supplies, it must reduce wild fish inputs in feed and adopt more ecologically sound management practices.

The worldwide decline of ocean fisheries stocks has provided impetus for rapid growth in fish and shellfish farming, or aquaculture. Between 1987 and 1997, global production of farmed fish and shellfish (collectively called 'fish') more than doubled in weight and value, as did its contribution to world fish supplies¹. Fish produced from farming activities currently accounts for over one-quarter of all fish directly consumed by humans. As the human population continues to expand beyond 6 billion, its reliance on farmed fish production as an important source of protein will also increase.

Growth in aquaculture production is a mixed blessing, however, for the sustainability of ocean fisheries. For some types of aquaculture activity, including shrimp and salmon farming, potential damage to ocean and coastal resources through habitat destruction, waste disposal, exotic species and pathogen invasions, and large fish meal and fish oil requirements may further deplete wild fisheries stocks². For other aquaculture species, such as carp and molluscs, which are herbivorous or filter feeders, the net contribution to global fish supplies and food security is great³. The diversity of production systems leads to an underlying paradox: aquaculture is a possible solution, but also a contributing factor, to the collapse of fisheries stocks worldwide.

Here we examine marine and freshwater fish farming activities around the world and ask: does aquaculture enhance—or diminish—the available fish supply? This is an important scientific and policy issue, and one that also addresses the common perception that aquaculture is an 'add on' to current ocean fish productivity. Many people believe that aquaculture production will compensate for the shortfall in ocean harvests as ocean fisheries deteriorate, or that fish farming will restore wild populations by relieving pressure on capture fisheries. We conclude that the compensation argument is correct for some aquaculture practices but unfounded for others. We do not find evidence that supports the restoration argument.

Our analysis focuses on aquaculture trends in the past 10–15 years—a period of heightened ecological and economic integration between capture fisheries and aquaculture activities. We limit our discussion to finfish, bivalves and crustaceans, which collectively make up three-quarters of global aquaculture production by weight, and exclude seaweed production¹. Ocean fisheries and aquaculture

now share or compete for many coastal ecosystem services, including the provision of habitat and nursery areas, feed and seed (larvae) supplies, and assimilation of waste products. Aquaculture and ocean fisheries are further linked economically through competition in world markets for the sale of their products, and biologically through exotic species invasions and pathogen transmission. Each of these connections is examined below.

As aquaculture production continues to increase and intensify, both its reliance and its impact on ocean fisheries are likely to expand even further. The balance between farmed and wild-caught fish, as well as the total supply of fish available for human consumption, will depend on future aquaculture practices. In the final section, we explore technological, management and policy options for sustaining aquaculture production. We argue that farming can contribute to global (net) fish supplies only if current trends in fish meal and fish oil use for aquaculture are reversed and policies are enforced to protect coastal areas from environmental degradation.

Aquaculture is a diverse activity

More than 220 species of finfish and shellfish are farmed; the range includes giant clams, which obtain most of their nutrients from symbiotic algae; mussels, which filter plankton; carps, which are mainly herbivorous; and salmon, which are carnivorous¹. Two key criteria, ownership of stock and deliberate intervention in the production cycle (husbandry), distinguish aquaculture from capture fisheries. Fish farming typically involves the enclosure of fish in a secure system under conditions in which they can thrive. Interventions in fish life cycles range from exclusion of predators and control of competitors (extensive aquaculture) to enhancement of food supply (semi-intensive) to the provision of all nutritional requirements (intensive). Intensification implies increasing the density of individuals, which requires greater use and management of inputs, greater generation of waste products and increased potential for the spread of pathogens.

Production practices and their impacts on marine ecosystems vary widely. Molluscs are generally farmed along coastlines where wild or hatchery-reared seed are grown on the seabed bottom or in suspended nets, ropes or other structures. The animals rely entirely

on ambient supplies of plankton and organic particles for food. Several systems—ponds, tanks or cages—are used in farming finfish. Most marine and diadromous finfish are reared in floating net cages nearshore, and all their nutrition is supplied by formulated feeds. Carp and other freshwater finfish are usually grown in ponds, often integrated within agricultural ecosystems. Shrimp dominate crustacean farming and are grown in coastal ponds. Farming of both shrimp and freshwater finfish varies in its intensity and dependence on formulated feeds.

Within aquaculture's wide diversity of species and production practices, two distinct subsectors have emerged during the past decade⁴. The first group includes commercial farms that primarily use intensive and semi-intensive methods to produce medium- to high-valued commodities for regional or global markets. The other group encompasses family and cooperative farms that rely on extensive and semi-intensive practices to produce low-value species for household subsistence or local markets. Some divisions between these sectors are becoming blurred. In China and other parts of Asia, for example, many small-scale farming operations are intensifying as land and water resources become increasingly scarce and valuable⁵.

Harvested weight and value for some of the most widely consumed aquaculture species are shown in Table 1. Asia accounts for roughly 90% of global aquaculture production, and China alone contributes more than two-thirds of the total¹. Europe, North America and Japan collectively produce just over one-tenth of global aquaculture output but consume the bulk of farmed seafood that is traded internationally.

The production of carp has increased markedly in Asia (mainly China) for local or regional consumption by relatively low-income households. In contrast, increased volumes of salmon, shrimp and other high-value species have been marketed mainly in industrialized countries. Farmed output and markets for other lower-value species, such as tilapia and milkfish, have increased in both developing and industrialized countries. Most farmed molluscs are still consumed locally and regionally in China and in other developing countries. However, production for global markets of certain species, including the Pacific cupped oyster, blue mussel, New Zealand mussel and Yesso scallop, has increased in several developed countries¹.

Market dynamics affecting both the supply and demand for aquaculture products differ sharply among types of fish. Expanding aquaculture production can alleviate pressure on wild fisheries stocks; for example, increasing the production of farmed fish that compete directly with wild fish (such as shrimp, salmon and molluscs) reduces prices and creates conditions that can lower investments in fishing fleets and fishing effort over time. Other farmed fish, such as tilapia, milkfish and channel catfish, provide alternatives to ocean fish such as cod, hake, haddock and pollock. Because niche markets have started to develop for several types of wild-caught fish, however, capture rates have remained high even as the production of viable substitutes has increased⁴.

The ability of the aquaculture sector to replace or provide market

alternatives for ocean catches depends significantly on the economics and policies of fisheries. High fixed costs of fishing fleets, inelastic supplies of labour in the fishing industry, and continued subsidies to the fisheries sector that approach 20–25% of gross revenue globally⁶ may mean that increased aquaculture production will not result in lower catches of wild fish in the short term. In the case of salmon, increased farm production has not resulted in reduced capture levels despite 30–50% declines in international prices for four of the five main species of wild salmon (chinook, coho, pink and chum) during the 1990s. Salmon catches worldwide actually rose by 27% between 1988 and 1997 (ref. 7). Similarly, despite rapid growth in alternative farmed fish like tilapia, wild capture of hake and haddock remained relatively stable during the past decade⁸. These examples show little obvious effect of aquaculture production on capture rates of wild fish.

Fishing down and farming up the food web

Capture fisheries landings as a whole have plateaued at around 85–95 Mt (million metric tonnes, or megatonnes) per year⁸. Moreover, there has been a gradual shift in wild fish capture from large and valuable carnivorous species to smaller, less valuable species that feed at lower trophic levels⁹. Although catch rates for some species have not declined during the 1990s, most ocean fisheries stocks are now recognized as over or fully fished¹⁰.

Aquaculture production, meanwhile, has surged, particularly during the past 10–15 years. Farmed fish supplies totalled 29 Mt in 1997, compared with 10 Mt a decade ago¹. Such growth helps to explain current patterns in ocean fish capture; between 1986 and 1997, 4 of the top 5, and 8 of the top 20 capture species were used in feed production for the aquaculture and livestock industries⁸. These species—anchoveta, Chilean jack mackerel, Atlantic herring, chub mackerel, Japanese anchovy, round sardinella, Atlantic mackerel and European anchovy—are all small pelagic fishes.

Many intensive and semi-intensive aquaculture systems use 2–5 times more fish protein, in the form of fish meal, to feed the farmed species than is supplied by the farmed product¹¹. In contrast, extensive and traditional systems use little or no fish meal, although nutrient-rich materials are often added to the water to stimulate growth of algae and other organisms on which the fish feed. Dietary requirements vary widely among fish species, depending on the aquaculture system, fish meal source and other dietary components (Table 2).

About 80% of carp and 65% of tilapia worldwide are farmed without the use of modern compound feeds—feeds formulated from multiple ingredients. In China, farmed production of carp and other omnivorous species is intensifying, however, and new commercial feed mills are being developed to serve these industries^{5,12}. More modern, intensive systems for herbivorous and omnivorous finfish rely heavily on added feeds, because fish are stocked at high densities that cannot be supported by natural food sources. Such systems, for example, US catfish farms, generally use compound feeds that contain high percentages of protein supplements from soybean meal, cottonseed meal and peanut meal¹³. Compound feeds for herbivorous and omnivorous finfish can also contain low to moderate levels of protein from fish and terrestrial animals.

In contrast, fish meal and fish oil are dominant ingredients in compound feeds for carnivorous finfish and marine shrimp. These two ingredients supply essential amino acids (such as lysine and methionine) that are deficient in plant proteins and fatty acids (eicosapentanoic acid (EPA) and docosahexanoic acid (DHA)) not found in vegetable oils¹⁴. They also provide energy, which is important because fish tend to convert carbohydrates to energy inefficiently¹⁴.

Herbivorous, omnivorous and carnivorous finfish all require about the same quantity of dietary protein per unit weight. But herbivorous and omnivorous freshwater finfish, such as carp, utilize plant-based proteins and oils better than carnivorous finfish, and

Table 1 Global weight and value for nine of the most widely consumed aquaculture species^{1,87}

Species	1997 weight (kilotonnes)	Annual weight growth (1987–97) %	1997 value in US\$ (millions)
Common carp	2,237	7.6	2,709
Grass carp	2,662	15.9	2,444
Silver carp	3,146	7.8	2,917
Nile tilapia	742	18.0	885
Channel catfish	238	3.4	372
Atlantic salmon	639	22.4	2,113
Milkfish	393	1.7	697
Giant tiger prawn	490	10.6	3,501
Pacific cupped oyster*	2,968	9.5	3,164

* Weight includes shell.

they require minimal quantities of fish meal to supply essential amino acids¹⁴. Nevertheless, compound feeds for omnivorous fish, such as tilapia, often contain about 15% fish meal, exceeding required levels¹¹. Manufacturers often over-formulate feeds, in part because dietary information for particular species is insufficient.

Because of the high levels of fish meal and fish oil in aquaculture feeds, many species require more fish biomass as inputs than the farmed fish produced. For the ten types of fish most commonly farmed, an average of 1.9 kg of wild fish is required for every kilogram of fish raised on compound feeds (Table 2). Only three of the ten types of fish—catfish, milkfish and carp—require less fish as inputs than is ultimately harvested. (Marine molluscs and many filter-feeding carp are not fed compound feeds at all.) In contrast, carnivorous species require 2.5–5 times as much fish biomass as feed as is produced.

Although aquaculture has the fastest growing demand for fish meal and fish oil, fish are not the only animals fed diets containing fish meal. The poultry and swine industries are the world's largest consumers of fish meal¹⁵. The proportion of fish meal in aquaculture feeds is, however, much higher than in poultry and livestock feeds, which on average contain only 2–3% fish meal as a protein supplement. The production of a kilogram of pork or poultry typically uses large amounts of plant proteins, but only a few hundred grams of fish, whereas production of a kilogram of carnivorous fish can use up to 5 kg of wild fish¹⁶.

The relative feed efficiency of fish farming is a complex subject that has not yet been fully analysed. Some aquaculture proponents argue that even if farmed fish production requires more wild fish biomass than is ultimately harvested, it is still more efficient than the production of commercially valuable carnivorous species in the wild¹⁷. Assuming a canonical value of a 10% energy flow between trophic levels¹⁸, producing 1 unit of predatory fish requires 10 units of food (largely small pelagic fish) compared with 2–5 units to produce a unit of farmed fish. This comparison is subject to debate, because energy flows between marine fish at different trophic levels are not well documented. Nevertheless, such efficiency comparisons bolster the logic for using some small pelagic fish in fish feeds.

Regardless of the exact efficiency ratio used, however, the growing aquaculture industry cannot continue to rely on finite stocks of wild-caught fish, a number of which are already classified as fully exploited, overexploited or depleted^{8,10}. Taking efficiency arguments to their logical conclusion—that ever increasing amounts of small pelagic fish should be caught for use in aquaculture feeds to expand the total supply of commercially valuable fish—would clearly be disastrous for marine ecosystems. Such an approach would also

severely constrain the long-term growth of the aquaculture industry itself.

Appropriation of net aquatic primary production

Data in the preceding section indicate that feed requirements for some types of aquaculture systems place a strain on wild fish stocks. But what is the aggregate impact of fish farming on ocean fisheries and marine resources? Tracing the flow of net aquatic primary production that moves through aquaculture (Fig. 1) provides a framework for assessing whether farmed fish production adds to global fish supplies on a net basis.

An estimated 8% of total aquatic primary production (137,000 Mt dry weight) is needed to sustain capture fisheries, seaweed collection and aquaculture; this proportion ranges from 2% in the open ocean to 24–35% in freshwater, shelf and upwelling systems¹⁹. Global capture fisheries (plus aquatic plants) remove 123 Mt from the sea and lakes²⁰, of which 27 Mt is directly discarded as bycatch²¹.

Capture fisheries landings (excluding discarded bycatch) amount to 96 Mt, of which 65 Mt of whole fish and 1 Mt of seaweeds are consumed by humans. The remaining 30 Mt of fish catch plus another 2 Mt of processing scraps from aquaculture and fisheries are used for fish meal production²². The fish meal industry has proposed that fishing vessels be encouraged to retain bycatch, now discarded, for sale to producers of fish meal and fish oil¹⁵. Sale of bycatch could prove undesirable, however, if it undermines efforts to reduce bycatch rates or decreases in situ recycling of bycatch.

One-third of the fish used to make fish meal inputs, ~10 Mt, is converted to aquaculture feeds^{20,22}. The remaining two-thirds of the fish, ~22 Mt, is used to make fish meal for chicken, pig and other animal feeds, although the share of aquaculture continues to increase. The proportion of fish meal supplies used for farming fish rose from 10% in 1988 to 17% in 1994 and 33% in 1997 (refs 22–24).

Other feed inputs to aquaculture are derived from terrestrial agriculture or, in the case of filter-feeding molluscs, from planktonic production. Pelagic and benthic microalgae are also consumed directly by herbivorous and omnivorous carp and tilapias and are thus important in extensive and semi-intensive freshwater ponds common in the tropics. Mollusc farming and other extensive aquaculture do not use compound feeds and do not therefore appropriate fisheries production directly.

Total aquaculture production of finfish, crustaceans and molluscs amounts to 29 Mt (plus 8 Mt of farmed seaweeds). However, the net volume of fish flowing to human consumption through aquaculture is at maximum 19 Mt after ocean fisheries capture for fish feeds is

Table 2 Wild fish inputs used in feed for the ten types of fish and shell fish most commonly farmed in 1997*

Farmed fish	Total production (kilotonnes)	Percentage produced with compound feeds (by weight)	Production with compound feeds (kilotonnes)	Percentage fishmeal in feed	Percentage fish oil in feed	Average feed conversion ratio	Wild fish used for fishmeal (kilotonnes)	Ratio of wild fish: fed farmed fish†
Marine finfish‡	754	50	377	50	15	2.2	1,944	5.16
Eel	233	50	117	50	10	2	546	4.69
Marine Shrimp	942	77	725	30	2	2	2,040	2.81
Salmon	737	100	737	45	25	1.5	2,332	3.16
Trout	473	100	473	35	20	1.5	1,164	2.46
Tilapia	946	35	331	15	1	2	466	1.41
Milkfish	392	20	78	10	3	2	74	0.94
Catfish	428	82	351	10	3	1.8	296	0.84
Carp§								
Fed	6,985	35	2,445	8	1	2	1,834	0.75
Filter-feeding	5,189	0	0	—	—	—	—	—
Molluscs	7,321	0	0	—	—	—	—	—
Total	24,400		5,634				10,695	1.90

* Taken from refs 1, 16, 23 and A. Tacon, personal communication.

† Ratio is wild fish used for fishmeal to farmed fish produced using compound feeds. We assume a 5:1 conversion ratio of fish ('wet weights') to fishmeal and that one-sixteenth of fishmeal is obtained from processing byproducts²².

‡ Marine finfish (other than salmon, which is listed separately because it is diadromous and because of its market significance) include flounder, halibut, sole, cod, hake, haddock, redfish, seabass, congers, tuna, bonito and billfish.

§ Fed carp refers to carp species that are sometimes fed compound feeds. Filter-feeding carp are silver carp, bighead carp and catla.

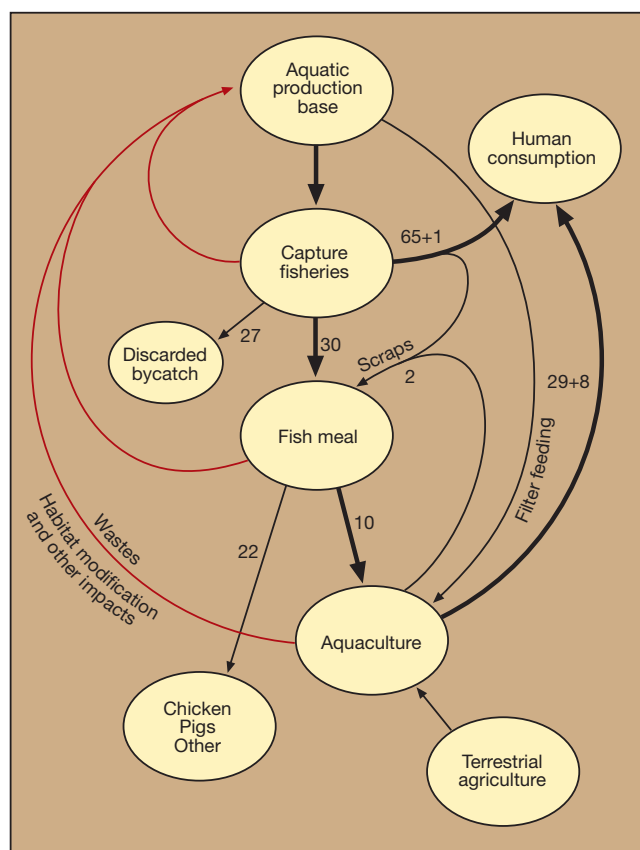


Figure 1 Flow chart of capture and farmed fisheries products from aquatic primary production. Numbers refer to 1997 data and are in units of megatons (million metric tonnes) of fish. Thicker lines refer to direct flows of aquatic primary production through

capture fisheries and aquaculture to humans. Thin lines refer to indirect and minor flows. Red lines indicate negative feedbacks on production base.

subtracted. The appropriation of aquatic productivity for fish feeds reduces supplies of wild fish that could potentially be consumed directly. In southeast Asia, for example, small pelagic fishes, such as mackerel, anchovy and sardines, provide an important protein source for people^{25,26}. Although some fish used for fishmeal and fish oil, such as menhaden, are distasteful to humans, the demand for small pelagic fish for direct human consumption is likely to increase with population growth in the developing world.

Ecological links between aquaculture and wild fish stocks

The use of wild fish to feed farmed fish places direct pressure on fisheries resources. But aquaculture can also diminish wild fisheries indirectly by habitat modification, collection of wild seedstock, food web interactions, introduction of exotic species and pathogens that harm wild fish populations, and nutrient pollution (Fig. 2). The magnitude of such effects varies considerably among aquaculture systems, but it can be great, as the following examples illustrate.

Habitat modification. Hundreds of thousands of hectares of mangroves and coastal wetlands have been transformed into milkfish and shrimp ponds. This transformation results in loss of essential ecosystem services generated by mangroves, including the provision of nursery habitat, coastal protection, flood control, sediment trapping and water treatment. Mangrove forests serve as nurseries that provide food and shelter to many juvenile finfish and shellfish caught as adults in coastal and offshore fisheries^{27–30}; in southeast Asia, mangrove-dependent species account for roughly one-third of yearly wild fish landings excluding trash fish³¹. A positive relationship between finfish and shrimp landings and mangrove area has been documented in Indonesia, Malaysia and the Philippines^{32–34}. Mangroves are also linked closely to habitat conditions of coral reefs and seagrass beds^{35,36}. Loss of mangrove

forests results in increased sediment transport onto downstream coral reefs. Fisheries capture from reefs contributes about 10% of human fish consumption globally and much more in developing countries³⁷.

The loss in wild fisheries stocks due to habitat conversion associated with shrimp farming is large. We estimate that a total of 400 g of fish and shrimp are lost from capture fisheries per kilogram of shrimp farmed in Thai shrimp ponds developed in mangroves (Box 1). If other fish and shellfish species caught in waterways adjoining mangrove areas are considered, the total reduction increases to 447 g of wild fish biomass per kilogram of shrimp raised. If the full range of ecological effects associated with mangrove conversion is accounted for, including reduced mollusc productivity in mangroves and losses to seagrass beds and coral reefs, the net yield from these shrimp farms is low—even without considering the use of fish meal in aquaculture feeds. Moreover, building aquaculture ponds in mangrove areas transforms fisheries from a common property resource available to multiple users to a privatized farm resource.

Use of wild seed to stock aquaculture ponds. Many aquaculture operations, especially extensive ponds, stock wild-caught rather than hatchery-reared finfish or shellfish post-larvae. Examples include farming of milkfish in the Philippines and Indonesia, tuna in South Australia, shrimp in south Asia and parts of Latin America and eels in Europe and Japan. In these systems, aquaculture is not a true alternative to wild harvests, but rather a means to raise wild fish to marketable size in captivity by lowering the high mortality rates characteristic of wild populations.

If bycatch rates are high, collecting seedstock for aquaculture operations can have very large consequences for wild fisheries. For example, milkfish constitute only 15% of total finfish fry collected

inshore by seine net³⁸—the remaining 85% are discarded and left to die on the beach. The 1.7 billion wild fry stocked annually in Philippine milkfish ponds³⁹ thus result in a loss of about 10 billion fry of other finfish species. In India and Bangladesh, up to 160 fish and shrimp fry are discarded for every fry of the giant tiger shrimp, *Penaeus monodon*, collected to stock shrimp ponds^{40,41}. The magnitude of annual fry bycatch is estimated at 62 million to 2.6 billion in three collecting centres in West Bengal, India⁴⁰.

Food web interactions. Many small pelagic fisheries exploited for feed are over-fished and are strained by climatic variability associated with El Niño Southern Oscillation events^{8,10}. The impact of pelagic fisheries depletion is thought to reduce available food supplies for marine predators, including valuable species consumed by humans^{42–44}. In the North Sea, for example, over-exploitation of many capelin, sandeel and Norway pout stocks, mainly for reduction to fishmeal, has been implicated in the declines of certain stocks of other wild fish such as cod^{9,45,46}, and changes in the distribution, populations sizes and reproductive success of various seal and seabird colonies^{47–49}. Similarly, a strong interaction between anchoveta and sea bird and mammal populations has been well documented for the Peruvian upwelling system⁵⁰.

Introduction of non-indigenous organisms. In some cases, aquaculture affects stocks of wild and farmed fish through biological pollution. Atlantic salmon—the dominant salmon species farmed—frequently escape from net pens. As much as 40% of Atlantic salmon caught by fishermen in areas of the North Atlantic Ocean are of farmed origin⁵¹. In the north Pacific Ocean, over 255,000 Atlantic salmon have reportedly escaped since the early 1980s and are caught by fishing vessels from Washington to Alaska⁵². Increasing evidence suggests that farm escapees may hybridize with and alter the genetic makeup of wild populations of Atlantic salmon which are genetically adapted to their natal spawning grounds⁵³. Such genetic alterations could exacerbate the decline in many locally endangered populations of wild Atlantic salmon^{53–55}.

Movement of stocks for aquaculture purposes can also increase the risk of spreading pathogens. The relationships between farmed and wild fish and disease transfer are complex and often difficult to disentangle. In Europe, however, serious epidemics of furunculosis and *Gyrodactylus salaris* in stocks of Atlantic salmon have been linked to movements of fish for aquaculture and re-stocking⁵⁶.

Since the early 1990s, Whitespot and Yellowhead viruses have caused catastrophic, multimillion-dollar crop losses in shrimp farms across Asia. Both pathogens have recently appeared in farmed and wild shrimp populations in the United States^{57,58} and the Whitespot virus has been reported in several countries in Central and South America (T. Flegel, personal communication; D. V. Lightner, personal communication). The Whitespot virus has

caused high mortalities in Texas shrimp farms⁵⁹ and may cause mortality of wild crustaceans (Joint Subcommittee on Aquaculture Virus Working Group, personal communication). This virus is thought to have been introduced into a Texas shrimp farm by release into nearby coastal waters of untreated wastes from plants processing imported Asian tiger shrimp⁶⁰, and by shipping of contaminated white shrimp *Litopenaeus vannamei* larvae throughout the Americas (T. Flegel, personal communication).

Effluent discharge. Untreated wastewater laden with uneaten feed and fish faeces may contribute to nutrient pollution near coastal ponds and cages^{61,62}. Pollution problems are most severe in shallow or confined water bodies⁶³; they also tend to be serious in regions where intensive aquaculture systems are concentrated. In many such areas, sedimentation of food particles and faecal pellets under and around fish pens and cages negatively affects the biogeochemistry of benthic communities⁶⁴. Moreover, nitrogen wastes (for example, ammonia and nitrite) that exceed the assimilative capacity of receiving waters lead to deterioration in water quality that is toxic to fish and shrimp⁶⁵. Problems of effluent discharge from aquaculture have been widely discussed, but management options for altering nitrogen biogeochemistry are based mostly on controlling the intensity of fish production in monoculture and polyculture systems⁶⁵. Aquaculturists have a stake in regulating nutrient pollution, because poor water quality and high stocking densities often promote outbreaks of pathogens and subsequent declines in farm productivity.

Towards sustainable aquaculture

The evidence presented above shows that total world aquaculture production currently adds to net global fish supplies, although many types of aquaculture result in a net loss of fish. Aquaculture's potential contribution to fish supplies is severely diminished by rapid growth in production of species fed carnivorous diets and by aquaculture practices that lead to coastal habitat destruction, biological pollution and discharge of untreated effluents. Continued expansion of aquaculture will require healthy coastal and freshwater ecosystems. Without clear recognition of the industry's dependence on natural ecosystems, it is unlikely that aquaculture will develop to its full potential or will continue to supplement ocean fisheries. We therefore suggest that the aquaculture industry prioritizes the following four chief goals: (1) expansion of the farming of low trophic level fish; (2) reduction of fish meal and fish oil inputs in feed; (3) development of integrated farming systems; and (4) promotion of environmentally sound aquaculture practices and resource management.

Farming down the food web. Carps and marine molluscs account for more than three-quarters of current global aquaculture output, and tilapia, milkfish and catfish contribute another 5% of total production. Fed mainly on herbivorous diets, these species provide most of the 19 Mt gain in fish supplies from aquaculture shown in Fig. 1. But market forces and government policies in many countries favour rapid expansion of high-value, carnivorous species, such as salmon and shrimp. Moreover, fish meal and fish oil are already being added to carp and tilapia feeds for weight gain, especially in Asia where farming systems are intensifying as a result of increased scarcity and value of land and freshwater resources. Given the huge volume of farmed carp and tilapia in Asia, significant increases in the fish meal and fish oil content of feed could place even more pressure on pelagic fisheries, resulting in higher feed prices and harm to marine ecosystems.

New initiatives by governments and international donor agencies are needed to further encourage farming of low trophic level fish with herbivorous diets^{66–69}. At the same time, more scientific research on the feed requirements of herbivores and omnivores is required to lessen the impetus to add fish meal and fish oil to their feeds⁶⁹. Substituting vegetable oils for fish oils in freshwater fish diets is technically possible—the n-3 fatty acids in fish oil are not

Box 1

Loss of wild fish from habitat conversion

For each hectare of mangrove, about 600 kg of finfish⁸⁸ and 600 kg of shrimp³⁴ are captured annually near shore in Malaysia. Applying this catch/mangrove area correlation to coastal regions of Thailand—where an estimated 65,000 ha of mangroves have been converted to shrimp ponds⁸⁹—indicates potentially significant losses in wild fish production. Shrimp ponds in Thailand have average annual yields of 3,000 kg ha⁻¹ (ref. 90); on balance, therefore, about 200 g each of fish and shrimp may be lost for every kilogram of shrimp farmed. Resident finfish species in mangrove waters comprise 97% of total fish biomass, which is equivalent to about 10.1 g m⁻² of mangroves⁹¹. These numbers translate into an estimated loss of more than 100 kg of on-site fish biomass for every hectare of converted mangrove, or 34 g resident fish per kilogram farmed shrimp. Adding capture fisheries losses within mangrove areas to those nearshore results in a total reduction of fisheries biomass of 434 g per kilogram farmed shrimp due to habitat conversion alone.

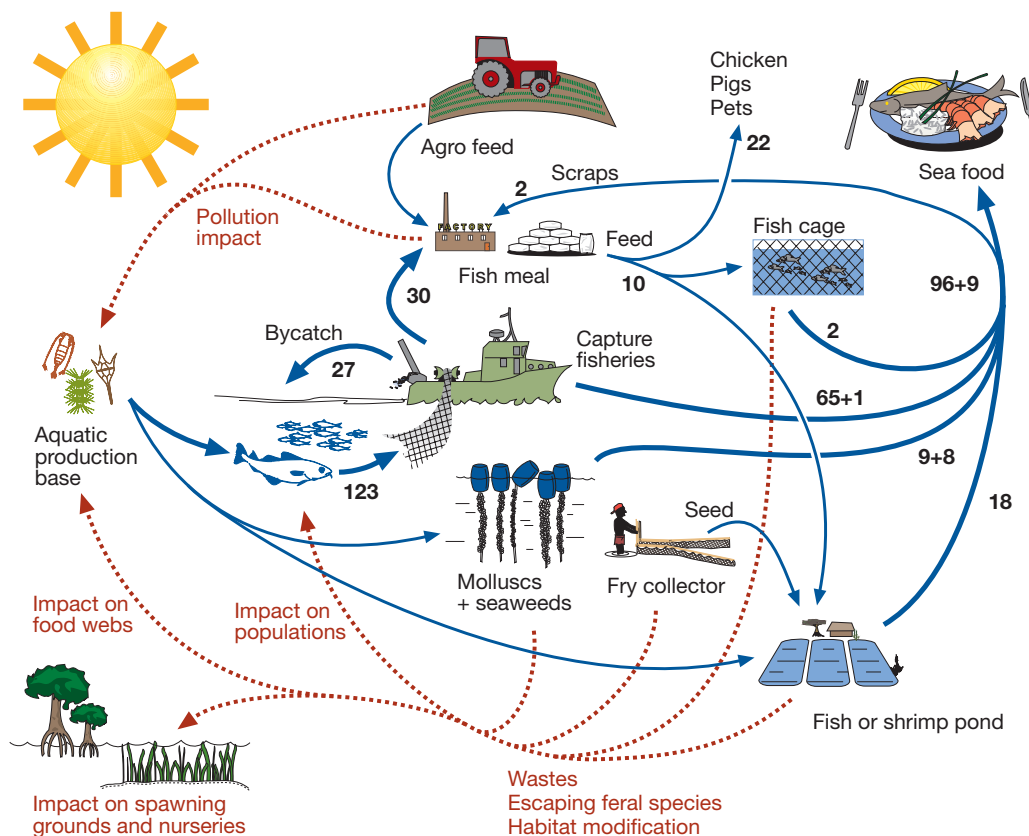


Figure 2 Ecological links between intensive fish and shrimp aquaculture and capture fisheries. Thick blue lines refer to main flows from aquatic production base through fisheries and aquaculture to human consumption of seafood. Numbers refer to 1997 data

and are in units of megatons (million metric tonnes) of fish, shellfish and seaweeds. Thin blue lines refer to other inputs needed for production. Hatched red lines indicate negative feedbacks.

essential in such diets⁷⁰—but the fatty acid profile and thus flavour and marketability may be affected^{71,72}. Moreover, some herbivorous fish appear to have more robust immune systems when fish oil is included in their diet⁷³.

Reducing fish meal and fish oil inputs in feed. Feed is the largest production cost for commercial aquaculture (for example, most farming of salmon, other marine finfish and shrimp), and thus improving feed efficiency in industrial systems is already a priority. Moreover, fish meal prices have risen in real terms in the past three decades and are likely to increase further with continued growth in demand. Increases in fish meal and fish oil prices could undermine the profitability of many aquaculture enterprises¹⁶.

Research to develop substitutes for these feed ingredients is now focused on commodities such as oilseeds (especially soybeans), meat byproducts (such as blood meal and bone meal) and microbial proteins^{74,75}. Already the fish meal content of some feeds has been reduced considerably, for example in the salmon industry, albeit largely by substitution with cheaper fish oil. Nevertheless, complete replacement of fish meal and fish oil in aquaculture feeds faces severe barriers. Especially for carnivorous fishes, vegetable proteins have inappropriate amino-acid balance and poor protein digestibility, although inclusion of meat byproducts can help overcome this problem^{74,76}.

Substitution of fish oil with cheaper vegetable oil in aquaculture feeds may also affect consumer demand, as evidence suggests that the ratio of n-6:n-3 fatty acids in human diets is already too high^{70,77}. There are, however, alternative sources of n-3 fatty acids for humans, including molluscs and other types of seafood, and research is underway to increase the n-3 fatty-acid content in poultry products and in oilseeds used for feed^{78,79} (W. F. Kirk, personal communication). A move towards partial substitution of

plant and terrestrial animal proteins for fish proteins in feed is widely accepted within the aquaculture industry, but the urgency of such efforts remains controversial. Because over-exploitation of pelagic fisheries has negative ecological and social consequences, developing a strategy to replace fish meal and fish oil in feeds should become both a private and public-sector priority.

Integrating production systems. Polyculture systems have been used for centuries. Even today, four of the most widely cultivated fish species are produced together in the same pond in China: silver carp (a phytoplankton filter feeder), grass carp (a herbivorous macrophyte feeder), common carp (an omnivorous detritus bottom feeder) and bighead carp (a zooplankton filter feeder)^{69,80}. This type of system efficiently utilizes available food resources and water resources (that is, surface, pelagic and benthic) of the pond ecosystem, with the consequent effects of reducing costs and increasing productivity.

Integrated systems can also be used for high-valued fish, such as salmon and shrimp, to reduce effluents, diversify products and increase productivity. Several studies show that seaweed and mussels grow well in wastewater from intensive and semi-intensive systems, thereby reducing nutrient and particulate loads to the environment^{81–85}. For example, in Chile salmon can be farmed with *Gracilaria chilensis* (an agarophytic red alga) that removes large amounts of dissolved nitrogen and phosphorous wastes from salmon cages⁸⁶. The effluent output from salmon farming is used to produce a seaweed crop, and the added revenue from the sale of the seaweed more than pays for the extra infrastructure needed for the integrated system. Policies that require producers to internalize environmental costs of effluent discharge (for example, through mandatory sewage treatment) can make such systems even more profitable. The marketability of molluscs raised in intensive fish

farming areas is currently constrained, however, by human health considerations that must be addressed to make these types of integrated system economically viable.

Promoting environmentally sound aquaculture and resource management. Long-term growth of the aquaculture industry requires both ecologically sound practices and sustainable resource management. Such practices can be encouraged by regulating the siting of ponds in mangroves and other coastal wetlands, establishing fines to minimize escapes from aquaculture netpens, enforcing strict biosafety measures for imported stock, and mandating treatment and recirculation of wastewater. Many aquaculturists have adopted such practices in the absence of strict policy measures, especially as environmental concerns have surfaced in recent years. In poor countries, however, these policies are often neither economically and socially feasible, nor politically enforceable. Despite significant improvements in the industry, there remains a considerable distance between ecologically sound technologies on the shelf and those actually implemented in the field. External funding agencies—such as development banks—are strategically positioned to influence the development of aquacultural technology, rehabilitation of ecosystems degraded by aquaculture and protection of coastal ecosystems.

How markets for resources are managed in the future will be a principal determinant of whether aquaculture depletes or enhances net fish supplies. The absence of regulations or price disincentives on coastal pollution, for example, limits mollusc farming and slows the adoption of non-polluting technologies by other marine aquaculture systems. Subsidies within the ocean fisheries sector often prevent farmed fish from substituting for wild fish catch, at least until fisheries are fully depleted.

Perhaps the largest unknown for both the private and public sectors is the future availability of freshwater for aquaculture production. Increasing scarcity of freshwater resources could severely limit the farming of herbivorous fish such as carps and tilapia. With a more binding constraint on freshwater systems, there is even more pressure to develop marine aquaculture systems that are ecologically and socially sound.

Mandate for the future

Fulfilling aquaculture's long-term potential to supplement global fish supplies and to provide food for the world's growing population will require a shared vision between the public and private sectors. Governments can support research and development on environmentally benign systems, eliminate implicit subsidies for ecologically unsound fish production, and establish and enforce regulatory measures to protect coastal ecosystems. At the same time, the private sector must alter its course and recognize that current practices that lead to dependence on pelagic fisheries, habitat destruction, water pollution and non-native introductions run counter to the industry's long-term health. If public and private interests act jointly to reduce external costs generated by farming systems, present trends may be reversed and the net contribution of aquaculture to global fish supplies can become increasingly positive. Without this shared vision, an expanded aquaculture industry poses a threat, not only to ocean fisheries, but also to itself. □

1. Food and Agricultural Organization *Aquaculture Production Statistics 1988–1997* (Food and Agricultural Organization, Rome, 1999).
2. Naylor, R. *et al.* Nature's subsidies to shrimp and salmon farming. *Science* **282**, 883–884 (1998).
3. Williams, M. in *Sustainable Aquaculture* (ed. Bardach, J. E.) 15–51 (Wiley, 1997).
4. Muir, J. F. & Young, J. A. Aquaculture and marine fisheries: will capture fisheries remain competitive? *J. Northw. Atl. Fish. Sci.* **23**, 157–174 (1998).
5. Cremer, M., Baoxin, Z., Schmittou, H. & Jian, Z. in *International Aquafeed Directory and Buyers' Guide 1999* 19–25 (Turret RAI, Middlesex, UK, 1999).
6. Milazzo, M. in *World Bank Technical Paper no. 406* (The World Bank, Washington DC, 1998).
7. Johnson, H. M. *Annual Report on the United States' Seafood Industry* 6th edn (H. M. Johnson & Associates, Bellevue, WA, 1998).
8. Food and Agricultural Organization *The State of World Fisheries and Aquaculture 1998* (Food and Agricultural Organization, Rome, 1999).

9. Pauly, D., Christensen, V., Dalsgaard, J., Froese, R. & Torres, F. Jr Fishing down marine webs. *Science* **279**, 860–863 (1998).
10. National Research Council *Sustaining Marine Fisheries* (National Academy Press, Washington DC, 1999).
11. Tacon, A. G. J. Feeding tomorrow's fish. *World Aquaculture* **27**, 20–32 (1996).
12. Crook, F. W., Hsu, H. & Lopez, M. in *Agricultural Outlook 13–16* (USDA Economic Research Service, Washington DC, 1999).
13. Robinson, E. Channel catfish. *Int. Aquafeed* **1**, 15–23 (1998).
14. De Silva, S. S. & Anderson, T. A. *Fish Nutrition in Aquaculture* (Chapman and Hall, London, 1994).
15. Pike, I. H. in *International Aquafeed Directory* (ed. Fraser, S.) 39–49 (Turret, Middlesex, UK, 1998).
16. Tacon, A. C. G. in *International Aquafeed Directory* (ed. Fraser, S.) 5–37 (Turret, Middlesex, UK, 1998).
17. Forster, J. Aquaculture chickens, salmon: a case study. *World Aquaculture Magazine* **30**, 33,35–38, 40,69–70 (1999).
18. Odum, E. P. *Fundamentals of Ecology* (W. B. Saunders, Philadelphia, 1979).
19. Pauly, D. & Christensen, V. Primary production required to sustain global fisheries. *Nature* **374**, 255–257 (1995).
20. Tacon, A. FAO aquaculture production update. *Int. Aquafeeds* **2**, 13–16 (1998).
21. Alverson, D. L., Freeberg, M. H., Murawski, S. A. & Pope, J. G. *FAO Fisheries Technical Paper no. 339* (Food and Agricultural Organization, Rome, 1994).
22. Pike, I. H. Fishmeal outlook. *Int. Aquafeeds* **1**, 5–8 (1998).
23. Tacon, A. Estimated global aquafeed production and aquaculture in 1997 and projected growth. *Int. Aquafeed* **2**, 5 (1999).
24. Barlow, S. in *Fish Farmer* 40–43 (Amben, Surrey, UK, 1989).
25. Dalzell, P. & Ganaden, R. A. *A Review of the Fisheries for Small Pelagics in Philippine Waters*. (Bureau of Fisheries and Aquatic Resources, International Center for Living Aquatic Resources Management, Quezon City, The Philippines, 1987).
26. Trinidad, A. C., Pomeroy, R. S., Corpuz, P. V. & Aguero, M. *Bioeconomics of the Philippine Small Pelagics Fishery* (International Center for Living Aquatic Resources Management, Manila, The Philippines, 1993).
27. Robertson, A. I. & Duke, N. C. Mangroves as nursery sites: comparisons of the abundance and species composition of fish and crustaceans in mangroves and other nearshore habitats in tropical Australia. *Mar. Biol.* **96**, 193–205 (1987).
28. Primavera, J. H. Fish predation on mangrove-associated penaeids: The role of structures and substrate. *J. Exp. Mar. Biol. Ecol.* **215**, 205–216 (1997).
29. Primavera, J. H. Mangroves as nurseries: shrimp populations in mangrove and non-mangrove habitats. *Est. Coast. Shelf Sci.* **46**, 457–464 (1998).
30. Ronnback, P. The ecological basis for economic value of seafood production supported by mangrove ecosystems. *Ecological Economics* **29**, 235–252 (1999).
31. Sasekumar, A., Chong, V. C. & Lim, K. H. in *Proceedings of the Third ASEAN-Australia Symposium on Living Coastal Resources: Status Reviews 1* (eds Wilkinson, C., Sudara, S. & Ming, C. L.) 139–165 (Australian Institute of Marine Science, Townsville, Australia, 1994).
32. Martosubroto P. & Naamin, M. Relationship between tidal forests (mangroves) and commercial shrimp production in Indonesia. *Mar. Res. Indonesia* **18**, 81–86 (1977).
33. Camacho, A. S. & Bagarinao, T. in *Mangroves of Asia and the Pacific: Status and Management, Technical Report UNDP/UNESCO Research and Pilot Programme on Mangrove Ecosystem in Asia and the Pacific (RAS/79/002)* (Natural Resources Management Center and National Mangrove Committee, Ministry of Natural Resources, Quezon City, The Philippines, 1987).
34. Sasekumar, A. & Chong, V. C. in *Proceedings of the 10th Annual Seminar of the Malaysian Society of Marine Sciences* (Sasekumar, A., Phang, S. M. & Chong, E. L.) 10–15 (Ardyas, Kuala Lumpur, Malaysia, 1987).
35. Ogden, J. C. The influence of adjacent systems on the structure and function of coral reefs. *Proceedings of the International Coral Reef Symp.* **1**, 123–129 (1988).
36. Yanez-Arancibia, A., Lara-Dominguez, A. L. & Day, J. W. Jr Interactions between mangroves and seagrass habitats mediated by estuarine nekton assemblages: coupling of primary and secondary production. *Hydrobiologia* **264**, 1–12 (1993).
37. Munro, J. J. in *Reef Fisheries* (eds Polunin, C. R. & Roberts, C. M.) 1–14 (Chapman and Hall, London, 1994).
38. Bagarinao, T. & Taki, Y. in *Indo-Pacific Fish Biology: Proceedings of the Second International Conference on Indo-Pacific Fishes* (eds Uyeno, T., Arai, R., Taniuchi, T. & Matsuura, K.) 728–739 (Ichthyological Society of Japan, Tokyo, 1986).
39. Bagarinao, T. in *Tropical Mariculture* (ed. de Silva, S.) 381–448 (Academic, London, 1998).
40. Banerjee, B. K. & Singh, H. *The Shrimp Bycatch in West Bengal* (Bay of Bengal Programme, Madras, India, 1993).
41. Islam, M. M., Rahman, S. L., Halder, G. C., Mazid, M. A. & Mahmood, N. Extent of damage to different crustaceans and fin fishes in collecting *Penaeus monodon* (Fabricius) post-larvae in Satkhira coastal region. *J. Mar. Biol. Assoc. India* **38**, 1–7 (1996).
42. Folke, C. & Kautsky, N. The role of ecosystems for a sustainable development of aquaculture. *Ambio* **18**, 234–243 (1989).
43. Hamer, K. C., Monaghan, P., Uttley, J. D., Walton, P. & Burns, M. D. The influence of food supply on the breeding ecology of kittiwakes (*Rissa tridactyla*) in Shetland. *Ibis* **135**, 255–263 (1993).
44. Fischer, J., Haedrich, R. L. & Sinclair, P. R. *Interecosystem Impacts of Forage Fish Fisheries* (Eco-Research Program, Memorial University of Newfoundland, Newfoundland, 1997).
45. Hislop, J. R. G. Changes in North Sea gadoid stocks. *ICES J. Marine Science* **53**, 1146–1156 (1996).
46. Monaghan, P. Seabirds and sandeels—the conflict between exploitation and conservation in the northern North Sea. *Biodivers. Conserv.* **1**, 98–111 (1992).
47. Wright, P. J. & Begg, G. S. A spatial comparison of common guillemots and sandeels in Scottish waters. *ICES J. Mar. Sci.* **54**, 578–592 (1997).
48. Brown, E. G. & Pierce, G. J. Monthly variation in the diet of harbour seals in inshore waters along the southeast Shetland (UK) coastline. *Mar. Ecol. Prog. Ser.* **167**, 275–289 (1998).
49. Phillips, R. A., Caldwell, R. W. G. & Furness, R. W. The influence of food availability on the breeding effort and reproductive success of Arctic skuas (*Stercorarius parasiticus*). *Ibis* **138**, 410–419 (1996).
50. Pauly, D. in *The Peruvian Anchoveta and its Upwelling Ecosystem: Three Decades of Change* (eds Pauly, D. & Tsukayama, I.) 325–342 (ICLARM, Manila, The Philippines, 1987).
51. Hansen, P., Jacobsen, J. A. & Und, R. A. High numbers of farmed Atlantic salmon, *Salmo salar*,

- observed in oceanic waters north of the Faroe Islands. *Aquaculture Fisheries Management* **24**, 777–781 (1993).
52. McKinnell, S. & Thomson, A. J. Recent events concerning Atlantic salmon escapees in the Pacific. *ICES J. Mar. Sci.* **54**, 1221–1225 (1997).
53. Gross, M. R. One species with two biologies: Atlantic salmon (*Salmo salar*) in the wild and in aquaculture. *Can. J. Fish. Aquat. Sci.* **55** (Suppl.1), 1–14 (1998).
54. McGinnity, P. *et al.* Genetic impact of escaped farmed Atlantic salmon (*Salmo salar* L.) on native populations: use of DNA profiling to assess freshwater performance of wild, farmed, and hybrid progeny in a natural river environment. *ICES J. Mar. Sci.* **54**, 998–1008 (1997).
55. Slaney, T. L., Hyatt, K. D., Northcote, T. G. & Fielden, R. J. Status of anadromous salmon and trout in British Columbia and Yukon Fisheries. *Am. Fish. Soc.* **21**, 20–35 (1996).
56. McVicar, A. H. Disease and parasite implications of the coexistence of wild and cultured salmon populations. *ICES J. Mar. Sci.* **54**, 1093–1103 (1997).
57. Nadala, E. C. B. & Loh, P. C. A comparative study of three different isolates of white spot virus. *Dis. Aquat. Organisms* **33**, 231–234 (1998).
58. Lightner, D. V. *A Handbook of Shrimp Pathological and Diagnostic Procedures for Diseases of Cultured Penaeid Shrimp* (World Aquaculture Society, Baton Rouge, Louisiana, 1996).
59. Nunan, L. M., Poulos, B. T. & Lightner, D. V. The determination of White Spot Shrimp Virus (WSSV) and Yellow Head Virus (YHV) in imported commodity shrimp. *Aquaculture* **160**, 19–30 (1998).
60. Lightner, D. V. *et al.* Risk of spread of penaeid shrimp viruses in the Americas by the international movement of live and frozen shrimp. *Rev. Sci. Tech. Off. Int. Epiz.* **16**, 146–160 (1997).
61. Braaten, B., Aure, J., Ervik, A. & Bøge, E. Pollution problems in Norwegian fish farming. *ICES Coastal Management* **26**, 11 (1983).
62. Gowen, R. J. & Bradbury, N. B. The ecological impact of salmonid farming in coastal waters: a review. *Oceanogr. Mar. Biol. Ann. Rev.* **25**, 563–575 (1987).
63. Iwama, G. K. Interactions between aquaculture and the environment. *Crit. Rev. Environ. Contr.* **21**, 177–216 (1991).
64. Ervik, A. *et al.* Regulating the local environmental impact of intensive marine fish farming. *Aquaculture* **158**, 85–94 (1997).
65. Hargreaves, J. A. Nitrogen biogeochemistry of aquaculture ponds. *Aquaculture* **166**, 181–212 (1998).
66. Tacon, A. J. Contribution to food fish supplies. *FAO Fisheries Circular* **886** (Rev. 1), 17–21 (1997).
67. Baily, C. Aquaculture and basic human needs. *World Aquaculture* **27**, 28–31 (1997).
68. El-Sayed, A. M. Alternative dietary protein sources for farmed Tilapia, *Oreochromis* spp. *Aquaculture* **179**, 149–168 (1999).
69. Tacon, A. J. & De Silva, S. S. Feed preparation and feed management strategies within semi-intensive fish farming systems in the tropics. *Aquaculture* **151**, 379–404 (1997).
70. Bell, J. G. in *Biology of Farmed Fish* (eds Black, K. D. & Pickering, A. D.) 114–145 (Sheffield Academic Press, Sheffield, UK, 1998).
71. Morris, C. A., Haynes, K. C., Keeton, J. T., and Gatlin, D. M. Fish oil and dietary pelagics and flavor of channel catfish. *J. Food Sci.* **60**, 1225–1227 (1995).
72. Steffens, W. Effects of variation in essential fatty acids in fish feeds on nutritive value of freshwater fish for humans. *Aquaculture* **151**, 97–119 (1997).
73. Pilarczyk, A. Changes in specific carp immune reaction caused by addition of fish oil to pellets. *Aquaculture* **129**, 425–429 (1995).
74. Webster, C. D., Tiu, L. G., Margan, A. M. & Gannam, A. Effect of partial and total replacement of fishmeal on growth and body composition of sunshine bass, *Morone chrysops* X *M. saxatilis*, fed practical diets. *J. World Aquaculture Soc.* **30**, 443–453 (1999).
75. Tacon, A. G. J. *Feed Ingredients for Carnivorous Fish Species: Alternatives to Fishmeal and Other Fishery Resources* (Food and Agricultural Organization, Rome, 1994).
76. Lim, C. & Dominy, W. G. in *Proceedings of the Aquaculture Feed Processing and Nutrition Workshop* (eds Akiyama, D. M. & Tan, R. K. H.) 163–172 (American Soybean Association, Singapore, 1991).
77. Pond, C. *The Fats of Life* (Cambridge Univ. Press, Cambridge, UK, 1998).
78. Huttenbauer, S. *Entrepreneurship in Food Preservation and Nutriceuticals* Harvard Business School Case Study (Harvard Business School Publishing, Boston, Massachusetts, 1999).
79. Mooney, J. W. Lipid and flavour quality of stored breast meat from broilers fed marine algae. *J. Sci. Food Agriculture* **78**, 134–140 (1998).
80. Bardach, J. E., Ryther, J. H. & McLarney, W. O. *Aquaculture: the Farming and Husbandry of Freshwater and Marine Organisms* (John Wiley, New York, 1972).
81. Jones, T. O. & Iwama, G. K. Polyculture of the Pacific oyster, *Crassostrea gigas* (Thunberg), with chinook salmon, *Oncorhynchus tshawytscha*. *Aquaculture* **92**, 313–322 (1991).
82. Neori, A. *et al.* Seaweed biofilters as regulators of water quality in integrated fish-seaweed culture units. *Aquaculture* **141**, 183–199 (1996).
83. Phang, S. -M., Shaharuddin, S., Noraishah, H. & Sasekumar, A. Studies on *Gracilaria changii* (Gracilariales, Rhodophyta) from Malaysian mangroves. *Hydrobiologia* **326/327**, 347–352 (1996).
84. Troell, M., Ronnback, P., Kautsky, N., Halling, C. & Buschmann, A. Ecological engineering in aquaculture: the use of seaweeds for removing nutrients from intensive mariculture. *J. Appl. Phycology* **11**, 89–97 (1999).
85. Soto, D. & Mena, G. Filter feeding by the freshwater mussel, *Diplodon chilensis*, as a biocontrol of salmon farming eutrophication. *Aquaculture* **171**, 65–81 (1999).
86. Troell, M. *et al.* Integrated marine cultivation of *Gracilaria chilensis* (Gracilariales, Rhodophyta) and salmon cages for reduced environmental impact and increased economic output. *Aquaculture* **156**, 45–61 (1997).
87. Food and Agricultural Organization Aquaculture production statistics 1987–1996. *Fish. Circ.* **815** (Rev. 10) (1998).
88. Singh, H. R., Chong, V. C., Sasekumar, A. & Lim, K. H. in *Proceedings of the Third ASEAN-Australia Symposium on Living Coastal Resources: Status Reviews* (eds Wilkinson, C., Sudara S. & Ming, C. L.) 105–122 (Australian Institute of Marine Science, Townsville, Australia, 1994).
89. Menasveta, P. Mangrove destruction and shrimp culture systems. *World Aquaculture* **28**, 36–42 (1997).
90. Rosenberry, B. (ed.) *World Shrimp Farming* (Shrimp News International, San Diego, California, 1998).
91. Ronnback, P., Troell, M., Primavera, J. H. & Kautsky, N. Distribution pattern of shrimps and fish among *Avicennia* and *Rhizophora* microhabitats in the Pagbilao mangroves, Philippines. *Est. Coast. Shelf. Sci.* **48**, 223–234 (1999).

Acknowledgements

We thank M. Williams, W. Falcon, V. Spruill, M. Drew, N. Wada, R. Kautsky, K. Jauncey, C. Tirado, R. Hoguet, R. Tatum and R. Mitchell for comments and assistance, and the David and Lucile Packard Foundation for funding.

Correspondence and requests for materials should be addressed to R.L.N. (e-mail: roz@leland.stanford.edu).

Deuterium in the Galactic Centre as a result of recent infall of low-metallicity gas

D. A. Lubowich^{*†}, Jay M. Pasachoff[‡], Thomas J. Balonek[§], T. J. Millar^{||}, Christy Tremonti[§], Helen Roberts^{||} & Robert P. Galloway[‡]

^{*} Department of Physics and Astronomy, Hofstra University, Hempstead, New York 11550, USA

[†] American Institute of Physics, Suite 1N01, 2 Huntington Quadrangle, Melville, New York 11747-4502, USA

[‡] Williams College–Hopkins Observatory, Williamstown, Massachusetts 01267, USA

[§] Department of Physics and Astronomy, Colgate University, Hamilton, New York 13346, USA

^{||} Department of Physics, University of Manchester Institute of Science and Technology, Box 88, Manchester M60 1QD, UK

The Galactic Centre is the most active and heavily processed region of the Milky Way, so it can be used as a stringent test for the abundance of deuterium (a sensitive indicator of conditions in the first 1,000 seconds in the life of the Universe). As deuterium is destroyed in stellar interiors, chemical evolution models¹ predict that its Galactic Centre abundance relative to hydrogen is $D/H = 5 \times 10^{-12}$, unless there is a continuous source of deuterium from relatively primordial (low-metallicity) gas. Here we report the detection of deuterium (in the molecule DCN) in a molecular cloud only 10 parsecs from the Galactic Centre. Our data, when combined with a model of molecular abundances, indicate that $D/H = (1.7 \pm 0.3) \times 10^{-6}$, five orders of magnitude larger than the predictions of evolutionary models with no continuous source of deuterium. The most probable explanation is recent infall of relatively unprocessed metal-poor gas into the Galactic Centre (at the rate inferred by Wakker²). Our measured D/H is nine times less than the local interstellar value, and the lowest D/H observed in the Galaxy. We conclude that the observed Galactic Centre deuterium is cosmological, with an abundance reduced by stellar processing and mixing, and that there is no significant Galactic source of deuterium.

The origin of deuterium has been extensively studied because it is not produced via stellar nucleosynthesis and any non-cosmological deuterium would be a signature of high-energy astrophysical processes. The D/H ratio is an important prediction of standard and non-homogeneous Big Bang models³ because the abundance of D depends critically on the temperature and baryonic density during the epoch of nucleosynthesis (the first 1,000 s) and might determine if the density is sufficient to close the Universe. However, D is completely destroyed in stellar interiors via $D(p,\gamma)^3\text{He}$ and not returned to the interstellar medium so that the abundance of D will decrease with time unless there are any additional sources of deuterium. Such sources, involving high-energy spallation reactions or large fluxes of protons or neutrons, would undermine the use of deuterium to estimate the baryonic density of the Universe and place constraints on Big Bang nucleosynthesis models.

The Galactic Centre is the most active and heavily processed region of the Galaxy⁴, and has a higher abundance of elements heavier than He (metallicity), faster star formation rate, and steeper initial mass function⁵. Thus the astration (recycling) rate in the Galactic Centre should be considerably larger than elsewhere in the Galaxy, resulting in a reduced D abundance. Chemical models at 12 Gyr of the Galactic bulge⁶ and the Galactic Centre predict the almost total destruction of deuterium giving $D/H = 3.2 \times 10^{-11}$ and $D/H = 5 \times 10^{-12}$, respectively. Thus if there were no additional sources of D, the Galactic Centre molecular clouds should be

composed primarily of astrated material completely depleted in D, and DCN should not be detectable. Thus the mere detection of D (in DCN) in the Sagittarius A molecular clouds requires a continuous source of deuterium to negate the effects of astration. Alternatively, if D is produced by any stellar or Galactic process, then it should be more abundant in the Galactic Centre and there should be a corresponding gradient in the D abundance⁷.

The largest Sgr A molecular clouds are the 50 km s^{-1} cloud (M-0.02-0.07) and the 20 km s^{-1} cloud (M-0.13-0.08), which are 10 pc from the Galactic Centre⁸. These are the appropriate objects in which to search for D because they are clearly related to the Galactic Centre activity. There is one reported marginal 1σ detection of deuterium from the DCN $J = 1-0$ line in the 50 km s^{-1} Sgr A molecular cloud core⁹ with corrected antenna temperature $\Delta T_a^* = 0.02 \pm 0.015 \text{ K}$, but no astrochemistry model was used and the D abundance was not determined. In this initial investigation we observed the 50 km s^{-1} cloud in May and June 1993, and February 2000, with the NRAO 12-m telescope at the position of the peak CS $J = 7-6$ and $J = 5-4$ emission¹⁰ (right ascension 17 h 42 min 42 s, declination $-28^\circ 58' 00''$), which ensured that we were observing the densest part of this cloud with an H_2 density $n(\text{H}_2) = (7.5 \pm 2.5) \times 10^5 \text{ cm}^{-3}$. Table 1 gives the results of our observations. We detected both the $J = 1-0$ and $J = 2-1$ lines of DCN, and obtained $T_R^* = 0.061 \pm 0.007 \text{ K}$ (8σ) and $T_R^* = 0.042 \pm 0.02 \text{ K}$

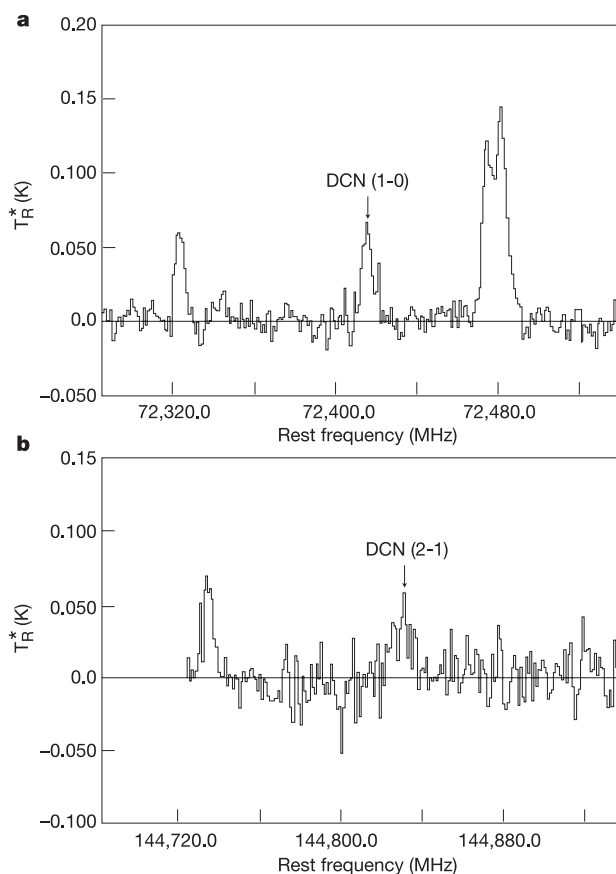


Figure 1 The DCN spectra in the Sgr A 50 km s^{-1} cloud. We used the NRAO 12-m telescope during 16–18 May 1993 and 29 June 1993 to observe the 50 km s^{-1} Sgr A Galactic Centre molecular cloud (M-0.02-0.07). We observed the DCN $J = 1-0$ and $J = 2-1$ lines at 72.4149 GHz and 144.83 GHz in total-power mode using position switching with the 3-mm and 2-mm SIS receivers. We used 1-MHz filters with a 256 MHz bandwidth in parallel, and obtained 4.14 km s^{-1} resolution with an 86-arcsec beam at 72 GHz and 2.07 km s^{-1} resolution with a 43-arcsec beam at 144 GHz. Pointing was checked using Jupiter and Uranus. The double sideband system temperature was 350 K at 72 GHz and 400 K at 144 GHz. **a**, The DCN $J = 1-0$ line. **b**, The DCN $J = 2-1$ line.

(2σ), respectively, where T_R^* is the source antenna temperature corrected for atmospheric attenuation and all telescope losses (ohmic and spillover) except for coupling of the source and beam. The DCN spectra are shown in Fig. 1.

The result of our unambiguous detection of DCN in the GC is *prima facie* evidence for the existence of Galactic Centre deuterium. Furthermore, we also mapped the distribution of $J = 1-0$ line of DCN at five additional positions offset by one arcmin from our centre position, and at one point offset by two arcmin south of our centre position, and detected DCN in four of these offset positions. We confirmed that the DCN $1-0$ line was optically thin by detecting the two strongest hyperfine components with the expected 5:3 ratio.

Because the HCN $J = 1-0$ line is optically thick, we used the optically thin $J = 1-0$ line of HC^{15}N to estimate $\text{DCN}/\text{HC}^{15}\text{N}$. To obtain the DCN/HCN abundance ratio, we followed the analysis of DCN/HCN in hot cores¹¹, which for a source extended relative to the size of the telescope beam gives

$$\text{DCN}/\text{HCN} = [T_R^*(\text{DCN})(\Delta V)/T_R^*(\text{HC}^{15}\text{N})(\Delta V)](^{15}\text{N}/^{14}\text{N})e^{(\Delta E/kT_{\text{ex}})}$$

where ΔE is the difference between the energies of the DCN and HC^{15}N transitions, and T_{ex} is the excitation temperature. For $T_{\text{ex}} = 75$ K, measured from 10 transitions of CH_3CCH in the Galactic Centre 50 km s^{-1} cloud¹², we obtain $\text{DCN}/\text{HC}^{15}\text{N} = 0.36 \pm 0.06$ and $\text{DCN}/\text{HCN} = (4.0 \pm 0.8) \times 10^{-4}$, using $(^{14}\text{N}/^{15}\text{N}) = 900$ measured for the Sgr A molecular cloud complex¹³. Our values are the smallest DCN/HCN and DCN/HCN ratios observed in any Galactic molecular cloud.

The observed molecular abundance ratio is, at low temperatures, always greater than the D/H ratio; this is because zero-point energy effects result in chemical fractionation, the magnitude of which depends on the molecule, chemistry and physical conditions, such as temperature and elemental abundances. To determine the degree of fractionation, $(\text{DCN}/\text{HCN})/(\text{D}/\text{H})$, expected for DCN (and other molecules) in Sgr A, we used a detailed time-dependent chemical kinetic model containing 5,260 gas-phase reactions,

Table 1 Molecular lines detected in the Sgr A 50 km s^{-1} molecular cloud

Molecule	Transition	ν (MHz)	T_R^* (mK)	r.m.s. (mK)	ΔV (km s^{-1})
Unknown		72,323.9	68	7	22.7
Unknown		72,344.2	18	7	30.6
DCN (centre)	1-0	72,414.9	61	7	30.9
DCN (hfs)	1-0	72,414.9	58	19	17.2
DCN (hfs)	1-0	72,413.5	38	19	26.2
Offset 1' E	1-0	72,414.9	40	19	33.1
Offset 1' W	1-0	72,414.9	30	16	41.5
Offset 1' S	1-0	72,414.9	55	16	33.9
Offset 2' S	1-0	72,414.9	32	15	17.8
HC^{13}CCN	8-7	72,475.1	120	7	21.9
HCC^{13}CN	8-7	72,482.1	145	7	27.9
HC^{15}N	1-0	86,055.0	192	27	26.8
SO	2(2)-1(1)	86,055.0	192	27	19.8
H^{13}CN	1-0	86,340.2	1,510	20	35.2
HCN	1-0	88,631.8	5,408	48	27.1
HCO^+		89,188.5	3,070	23	25.7
Unknown		89,204.3	532	23	11.7
Unknown		89,215.5	250	23	11.7
Unknown		89,221.8	130	23	7.8
HC^{13}CCN	10-9	90,593.1	130	23	18.5
HCC^{13}CN	10-9	90,601.8	120	23	25
HNC	1-0	90,663.5	2,198	23	41.3
C_2S	7,7-6,6	90,686.4	167	23	12.4
Unknown		144,735.3	67.6	15	12.6
DCN	2-1	144,828.0	38.9	15	25.2

The table gives the molecule, transition, rest frequency, peak gaussian line intensity T_R^* , and line widths (full-width at half-maximum, FWHM). We used the Lovas tables of rest frequencies²⁷ to identify the lines. The lines were well fitted by gaussian profiles. For broad lines not fitted by a gaussian we integrated over frequency while for overlapping lines we used several gaussians to fit the lines. DCN was detected at three offset positions. The centre position had the strongest DCN emission and the extended emission is similar to that observed in HCN and other molecules^{28,29}. The two hyperfine-structure lines of DCN were detected on 9 Feb 2000 with the MAC autocorrelator, a resolution of 195 kHz, and a system temperature of 295 K at 72 GHz. In the column headings, r.m.s. indicates the root mean square of the noise; ΔV is the line width of the spectral lines.

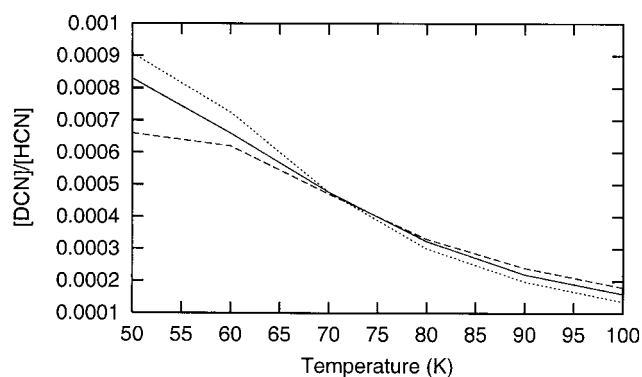


Figure 2 The DCN/HCN ratio calculated^{14,15} for $\text{D}/\text{H} = 1.7 \times 10^{-6}$ as a function of temperature. Data are shown for densities ($n(\text{H}_2)$) of $1 \times 10^5 \text{ cm}^{-3}$ (dashed line), $5 \times 10^5 \text{ cm}^{-3}$ (solid line), and $1 \times 10^6 \text{ cm}^{-3}$ (dotted line). Other physical conditions were chosen as appropriate for the Sgr A 50 km s^{-1} cloud (ionization rate, $1.3 \times 10^{-17} \text{ s}^{-1}$; and fractional abundances of C, N and O equal to 9.9×10^{-4} , 3.6×10^{-4} and 1.98×10^{-3} , respectively, approximately three times their solar abundances^{8,10,11,30}). At 75 K, the main routes to the formation of HCN and DCN are: $\text{HCO} + \text{N} \rightarrow \text{HCN} + \text{O}$; $\text{DCO} + \text{N} \rightarrow \text{DCN} + \text{O}$; $\text{C}_2\text{HD}^+ + \text{N} \rightarrow \text{CH}^+ + \text{DCN}$; $\text{C}_2\text{HD}^+ + \text{N} \rightarrow \text{CD}^+ + \text{HCN}$ where D is extracted from its reservoir, HD, and then transferred to other species.

165 non-deuterated species, 122 deuterated species, updated rate coefficients for the dissociative recombination of H_2D^+ , grain surface formation of H_2 and HD, and S chemistry^{14,15}.

We obtain a degree of fractionation of about 235 and best fit to DCN/HCN for $\text{D}/\text{H} = 1.7 \pm 0.3 \times 10^{-6}$; this value of D/H is nine times less than that in the local interstellar medium¹⁶ ($1.5 \pm 0.1 \times 10^{-5}$) but is 3.4×10^5 times larger than the value predicted by Galactic Centre models¹ (5×10^{-12}). Although the absolute abundances of DCN and HCN would be significantly changed by varying the physical conditions, our calculations show that the DCN/HCN ratio does not depend significantly on the density, metallicity and ionization rate (a faster ionization rate gave the same fractionation, but the system reached steady-state sooner). Figure 2 shows the DCN/HCN ratio as a function of temperature and density, and our calculations determined that the DCN/HCN ratio is almost constant, to within a factor of two, for the temperature range $65 \text{ K} < T < 85 \text{ K}$ and the density range $10^5 \text{ cm}^{-3} < n(\text{H}_2) < 10^8 \text{ cm}^{-3}$.

Our results are in agreement with the upper limits of $\text{D}/\text{H} < 1.2 \times 10^{-5}$ and $\text{D}/\text{H} < 8.3 \times 10^{-5}$ in the Galactic Centre 20 and 50 km s^{-1} clouds, respectively¹⁷, and $\text{DCN}/\text{HCN} < 6 \times 10^{-4}$ in the 50 km s^{-1} cloud¹⁸. Combining our result for the Galactic Centre $\text{D}/\text{H} = 1.7 \times 10^{-6}$ with the Sgr B2¹⁸ $\text{D}/\text{H} = 5 \times 10^{-6}$, the local interstellar medium $\text{D}/\text{H} = 1.5 \times 10^{-5}$, and the possible detection of D I with $\text{D}/\text{H} = 3.9 \times 10^{-5}$ in the Galactic anticentre interstellar clouds¹⁹, we obtain a positive abundance gradient of D/H in the Galaxy, indicating that there is no significant Galactic production of deuterium⁷.

The Galactic Centre deuterium comes from recent infall of metal-poor gas enriched in D but depleted in O (and other elements heavier than He) as compared to the values in the local interstellar medium. Our results confirm the theoretical models of the evolution of D in the Galactic disk by Prantzos²⁰, who analysed the D/O ratio and concluded that the disk D is the result of infall and astration. Thus D is not produced by the massive stars that produce O, which would yield a constant D/O ratio. Furthermore, our estimated D/H ratio for the Galactic Centre is consistent with the recent observations of infalling metal-poor gas from the halo or left over from the formation of the Galaxy (inflow of 1 solar mass per year of 0.1 solar metallicity gas)², whereas the resultant Galactic Centre D/H ratio is determined by the astration and mixing. This continued replenishment would negate much of the effect of astration.

Our results also constrain models of Galactic Centre activity. By comparing our results to models of activity in active galactic nuclei (AGN) where D is produced from cosmic-ray²¹ or γ -ray spallation reactions²², we conclude that the Galactic Centre has not had a recent AGN or quasar phase. Our results do not exclude weak activity—with a cosmic-ray proton luminosity $L_p = 8.5 \times 10^{41} \text{ erg s}^{-1}$ or γ -ray luminosity $L_\gamma = 1.7 \times 10^{40} \text{ erg s}^{-1}$ for 1 Gyr (which could produce our observed Galactic Centre D/H) coupled with periodic bursts of star formation—if the astration would reduce the Galactic Centre D/H to match the current deuterium abundance. However, these values of L_p and L_γ are respectively 42,500 times and 850 times larger than the current Galactic Centre cosmic-ray proton²³ or γ -ray²⁴ luminosities of $2 \times 10^{37} \text{ erg s}^{-1}$, but respectively 1,180 and 5,880 times less than quasar L_p or L_γ . Because almost all nucleosynthesis processes that can produce D over-produce Li or B by 10^3 – 10^5 times, the observed upper limit on the Galactic Centre (GC) Li of $(\text{Li}/\text{H})_{\text{GC}} < 3.9 \times 10^{-8}$ or $(\text{Li}/\text{H})_{\text{GC}} < 20 (\text{Li}/\text{H})_{\text{disk}}$ further implies there are no significant Galactic Centre sources of deuterium²⁵. (Here $(\text{Li}/\text{H})_{\text{disk}}$ is the lithium abundance in the Galactic disk.)

If all the deuterium is primordial and the astration models of Prantzos²⁰ are correct, then the primordial or early Galactic D/H = 5×10^{-5} . For this D/H, standard Big Bang nucleosynthesis models imply that the baryon-to-photon ratio is 3×10^{-10} , that there are fewer than four neutrino families, and that the baryon density of the Universe $\rho_b = 3 \times 10^{-31} \text{ g cm}^{-3}$ is less than the critical density $\rho_c = 3H_0^2/8\pi G = 9.2 \times 10^{-30} \text{ g cm}^{-3}$ (for a Hubble constant $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$) necessary to close the Universe assuming a Friedmann–Robertson–Walker cosmological model with $\Omega_b = \rho_b/\rho_c = 0.03$ (ref. 26). Thus the fraction of the critical density contributed by baryons (Ω_b) requires most of the baryons to be in the form of dark matter, and most of this dark matter to be non-baryonic³. □

Received 22 February; accepted 5 May 2000.

1. Audouze, J., Lequeux, J., Reeves, H. & Vignoux, L. Implications of the presence of deuterium in the galactic centre. *Astrophys. J.* **208**, L51–L54 (1976).
2. Wakker, B. P. *et al.* Accretion of low-metallicity gas by the Milky Way. *Nature* **402**, 388–390 (1999).
3. Schramm, D. N. & Turner, M. S. Big-bang nucleosynthesis enters the precision era. *Rev. Mod. Phys.* **70**, 303–318 (1988).
4. Yusef-Zadeh, F., Melia, F. & Wardle, M. The Galactic Center: an interacting system of unusual sources. *Science* **287**, 85–91 (2000).
5. Morris, M. Massive star formation near the Galactic center and the fate of the stellar remnants. *Astrophys. J.* **408**, 496–506 (1993).
6. Matteucci, F., Romano, D. & Molaro, P. Light and heavy elements in the galactic bulge. *Astron. Astrophys.* **341**, 458–468 (1999).
7. Pasachoff, J. M. & Vidal-Madjar, A. The need to observe the distribution of interstellar deuterium. *Comments Astrophys.* **14**, 61–68 (1989).
8. Poglitsch, A. *et al.* A survey of the 158 micron [C II] fine-structure line in the central 50 parsecs of the galaxy. *L33–L36* (1991).
9. Penzias, A. A. Interstellar HCN, HCO⁺, and the galactic deuterium gradient. *Astrophys. J.* **228**, 430–434 (1979).
10. Serabyn, E., Lacy, J. H. & Actermann, J. M. The compression of the M-0.02-0.07 molecular cloud by the Sagittarius A East shell source. *Astrophys. J.* **395**, 166–173 (1992).
11. Hatchell, J., Millar, T. J. & Rodgers, S. D. The DCN/HCN abundance ratio in hot molecular cores. *Astron. Astrophys.* **332**, 695–702 (1998).
12. Minh, Y. C., Irvine, W. M. & Friberg, P. Molecular abundances in the Sagittarius A molecular cloud. *Astron. Astrophys.* **258**, 489–494 (1992).
13. Güsten, R. & Ungerechts, H. Constraints on the sites of nitrogen nucleosynthesis from ¹⁵NH₃-observations. *Astron. Astrophys.* **145**, 241–250 (1985).
14. Rodgers, S. D. & Millar, T. J. The chemistry of deuterium in hot molecular cores. *Mon. Not. R. Astron. Soc.* **280**, 1046–1054 (1996).
15. Millar, T. J., Roberts, H., Markwick, A. J. & Charnley, S. B. The role of H₂D⁺ in the deuteration of interstellar molecules. *Phil. Trans. R. Soc. Lond. A* (in the press).
16. Linsky, J. L. Deuterium abundance in the local ISM and possible spatial variations. *Space Sci. Rev.* **84**, 285–296 (1998).
17. Lubowich, D. A., Anantharamaiah, K. R. & Pasachoff, J. M. A search for localized sources of noncosmological deuterium near the Galactic center. *Astrophys. J.* **345**, 770–775 (1989).
18. Jacq, T., Baudry, A., Walmsley, C. M. & Caselli, P. Deuterium in the Sagittarius B2 and Sagittarius A galactic center regions. *Astron. Astrophys.* **347**, 957–966 (1999).
19. Chengalur, J. N., Braun, R. & Butler, W. B. DI in the outer Galaxy. *Astron. Astrophys.* **318**, L35–L38 (1997).
20. Prantzos, N. The evolution of D and ³He in the Galactic disk. *Astron. Astrophys.* **310**, 106–114 (1996).
21. Ozernoi, L. & Chernomordik, V. V. The production of deuterium and helium-3 in the active galactic nucleus. *Sov. Astron.* **19**, 693–698 (1975).
22. Boyd, R. N., Ferland, G. J. & Schramm, D. N. Photoerosion and the abundances of the light elements.

Astrophys. J. **336**, L1–L4 (1989).

23. Mastichades, A. & Ozernoi, L. M. X-ray and gamma-ray emission of Sagittarius A* as a wind-accreting black hole. *Astrophys. J.* **426**, 599–603 (1994).
24. Mayer-Hasselwander, H. A. *et al.* High-energy gamma-ray emission from the Galactic Center. *Astron. Astrophys.* **335**, 161–172 (1998).
25. Lubowich, D. A., Turner, B. E. & Hobbs, L. M. Constraints on galactic center activity: a search for enhanced galactic center lithium and boron. *Astrophys. J.* **508**, 729–735 (1988).
26. Copi, C. J., Schramm, D. N. & Turner, M. S. Assessing Big-Bang nucleosynthesis. *Phys. Rev. Lett.* **75**, 3981–3984 (1995).
27. Lovas, F. J. Recommended rest frequencies for observed interstellar molecular microwave transitions—1991 revision. *J. Phys. Chem. Ref. Data* **21**, 181–272 (1992).
28. Fukui, Y. *et al.* HCN emission in the Sagittarius A molecular cloud. *Publ. Astron. Soc. Jpn* **29**, 643–667 (1977).
29. Genzel, R. *et al.* Far-infrared, submillimeter, and millimeter spectroscopy of the Galactic center—radio ARC and +20/+50 kilometer per second clouds. *Astrophys. J.* **356**, 160–173 (1990).
30. Simpson, J. P., Colgan, S. W. J., Rubin, R. H., Erickson, E. F. & Haas, M. R. Far-infrared lines from H II regions: Abundance variations in the galaxy. *Astrophys. J.* **444**, 721–738 (1995).

Acknowledgements

We thank H. Reeves and D. Tytler for comments, and A. Mancuso, S. Diaz, M. Pickard, R. Souza, K. Pagliuca and M. L. Kutner for their help. T.J.B. was a visiting scientist at the National Radio Astronomy Observatory, Tucson. The NRAO is a facility of the NSF operated under cooperative agreement by Associated Universities, Inc. We acknowledge a Research and Development Grant from Hofstra University, a Bronfman Science Center Grant from Williams College, and a PPARC grant at UMIST. J.M.P. and T.J.B. benefited from the Keck Northeast Astronomy Consortium.

Correspondence and requests for materials should be addressed to D.A.L. (e-mail: dlubowich@aip.org).

The mean free path for electron conduction in metallic fullerenes

O. Gunnarsson & J. E. Han

Max-Planck-Institut für Festkörperforschung, D-70506 Stuttgart, Germany

The electrical resistivity, ρ , of a metal is usually interpreted in terms of the mean free path (the average distance, l , an electron travels before it is scattered). As the temperature is raised, the resistivity increases and the apparent mean free path is correspondingly reduced. In this semi-classical picture, the mean free path cannot be much shorter than the distance, d , between two atoms. This has been confirmed for many systems and was considered to be a universal behaviour^{1,2}. Recently, some apparent exceptions were found, including alkali-doped fullerenes^{3–7} and high-temperature superconductors. However, there remains the possibility that these systems are in exotic states, with only a small fraction of the conduction electrons contributing to the conductivity; the mean free path would then have to be correspondingly larger to explain the observed resistivity. Here we report a model calculation of electron conduction in alkali-doped fullerenes, in which the electrons are scattered by intramolecular vibrations. The resistivity at large temperatures implies $l \ll d$, demonstrating that there is no fundamental principle requiring $l \gtrsim d$. At high temperatures, the semi-classical picture breaks down, and the electrons cannot be described as quasiparticles.

We have also calculated the resistivity due to electron–electron scattering for a half-filled Hubbard model. In this case the resistivity saturates and l is not very much smaller than d . This difference is traced to the difference between bosons and fermions. The resistivity is often calculated using the Boltzmann equation. Although this equation is usually derived semi-classically, assuming $l \gg d$, in our model for electron-vibration scattering it does not break down qualitatively at large temperature, T , where $l \ll d$. For small T the calculated ρ due to electron-vibration scattering has a linear dependence on T and a strong dependence on the pressure, in agreement with experiment⁸.

For A_3C_{60} ($A = K, Rb$) at $T \approx 500$ K, the resistivity is $\rho \approx 2\text{--}5$ m Ω cm. Assuming a spherical Fermi surface and three conduction electrons per site³, we find that $l \approx 1\text{--}2$ Å is almost an order of magnitude smaller than the separation $d = 10$ Å of the C_{60} molecules. The experimental data have substantial uncertainties, but this is unlikely to influence the qualitative discussion of l . Different experimental methods (direct and optical) for different types of samples (thin films and doped single crystals) all suggest that $l \ll d$ for large T . It has been suggested that the large ρ is due to the electrons being scattered inside the molecules. This would imply scattering between bands separated by at least 1 eV, which plays a fairly small role for experimental T . In addition, we find that this interband scattering² reduces the resistivity by providing an additional channel for conduction.

The conduction in A_3C_{60} takes place in a partly filled t_{1u} band. The T -dependent part of the resistivity is assumed to be due to scattering against phonons (vibrations) with H_g symmetry. We therefore consider a model with a threefold degenerate t_{1u} level and a fivefold degenerate H_g Jahn–Teller phonon on each molecule, the hopping between the molecules and the coupling between the electrons and the phonons. The hopping takes into account^{9,10} the orientational disorder of the molecules¹¹. The one-particle band width is $W = 0.6$ eV and the phonon frequency is ω_{ph} . The coupling is determined by the dimensionless $\lambda = (5/3)N(0)g^2/\omega_{ph}$, where $N(0)$ is the density of states per spin and g is the electron–phonon coupling constant.

We perform a finite-temperature quantum Monte Carlo (QMC) calculation¹², treating the phonons quantum mechanically. We calculate the current–current correlation function for imaginary times and make a transformation to real frequencies, using a maximum entropy method¹³. The QMC method has no ‘sign-problem’, and the resistivity of the model can be calculated essentially exactly down to quite small T .

Figure 1 shows the resistivity for a cluster of 48 C_{60} molecules for $\lambda = 0.5$ and $\omega_{ph} = 0.2$ eV. The QMC calculation (full line) shows that the resistivity can become very large, corresponding to $l \approx 0.7$ Å at $T = 0.5$ eV. By also considering an unrealistically large T , we emphasise the lack of a limitation of the type $l \gtrsim d$. Then there cannot be any general principle requiring $l \gtrsim d$, because such a principle would then also apply to the present model. Qualitatively similar results were obtained by Millis *et al.*¹⁴, using the dynamical mean-field theory (DMFT)¹⁵ and assuming classical phonons. Their calculation, however, does not prove that $l \ll d$ is possible, since it involves approximations. The moderate differences to our results are probably due to different models and their use of classical phonons.

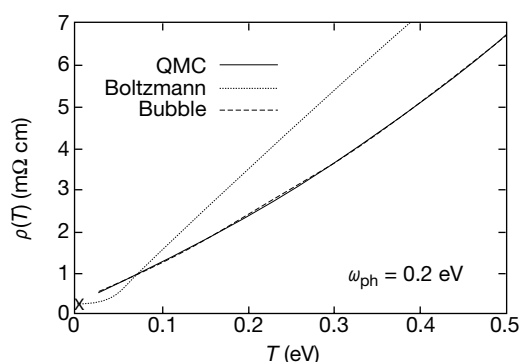


Figure 1 The resistivity ρ as a function of T . Results of the full QMC calculation, the Boltzmann equation (Bloch–Grüneisen) and the bubble diagram are shown. The cross shows the $\rho(T = 0)$ due to the orientational disorder. The figure illustrates that ρ can become extremely large, that the bubble calculation is quite accurate and that there is no qualitative breakdown of the Boltzmann equation at high T .

While the QMC calculation is essentially exact, it is difficult to interpret the results. For this purpose we use a diagrammatic approach. The idea is to simplify the diagrams as far as possible without losing the qualitative agreement with the QMC results. These diagrams are then studied to obtain a qualitative understanding of the results. This approach (Kubo formalism) requires the calculation of a bubble diagram including vertex corrections (see Fig. 2a). We neglect the vertex and calculate the bubble using the electron Green’s function from the QMC calculation. The resulting resistivity (dashed line in Fig. 1) is practically identical to the QMC result, justifying the neglect of vertex corrections for the present model. It was shown by Holstein¹⁶ that in the limit of a broad electronic band, all vertex corrections except ladder diagrams can be neglected and that a Boltzmann equation can be derived. Holstein’s derivation is not valid for the narrow band considered here, but our calculations show that his arguments are still helpful. For our model with a $\mathbf{q}$ -independent electron–phonon coupling, even the ladder diagrams can be neglected. By essentially following Holstein we obtain approximately a Boltzmann-like conductivity

$$\sigma(T) \approx \int d\omega N(\omega) \left(-\frac{df(\omega)}{d\omega} \right) \frac{1}{\text{Im} \Sigma(\omega)} |j_k|_{\epsilon_k=\omega}^2 \quad (1)$$

where $N(\omega)$ is the density of states, f is the Fermi function, $\Sigma(\omega)$ is the electron self-energy and j_k is the current matrix element for a state with the label k and the energy ϵ_k . We interpret $\text{Im} \Sigma$ to be the inverse of the relaxation time. For a large T , $\text{Im} \Sigma$ becomes comparable to or larger than the one-particle bandwidth. Because $\text{Im} \Sigma$ is related to the inverse life-time of the quasi-particles, this means that the lifetimes become so short that the concept of a quasi-particle breaks down at large T .

The resistivity thus depends crucially on Σ . To understand its behaviour, we consider the diagram in Fig. 2b calculated with bare Green’s functions and for simplicity neglecting the orbital degeneracy

$$\Sigma^{(1)}(\mathbf{k}, \omega) = g^2 \sum_{\mathbf{q}} \left[\frac{n_B(\omega_{ph}) + 1 - f(\epsilon_q)}{\omega - \omega_{ph} - \epsilon_q} + \frac{n_B(\omega_{ph}) + f(\epsilon_q)}{\omega + \omega_{ph} - \epsilon_q} \right] \quad (2)$$

where

$$n_B(\omega_{ph}) = \frac{1}{e^{\omega_{ph}/T} - 1} \xrightarrow{T \rightarrow \infty} \frac{T}{\omega_{ph}} \quad (3)$$

is the Bose occupation number. For large T , n_B becomes large, leading to a large $\text{Im} \Sigma$, a small σ and a large ρ . The Bose nature of the phonons is therefore of crucial importance for our result.

This is further illustrated by comparing with the resistivity due to the electron–electron scattering. Using the DMFT we have calculated the resistivity for a non-degenerate Hubbard model on a Bethe lattice and with $d = 1$ Å. We focus on the half-filled case, which is relevant for A_3C_{60} , and we do not consider the case of a doped Mott insulator¹⁷. Figure 3 shows $\rho(T)$ for different values of the on-site Coulomb interaction U . For $U/W \leq 1$ the system is a metal and $\rho(T)$ grows with T , whereas for $U/W \geq 1.25$ it is an insulator and $\rho(T)$ decreases with T . The important observation is that in the

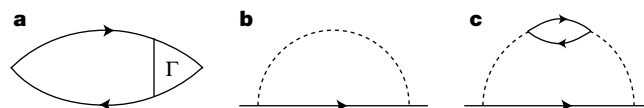


Figure 2 Relevant diagrams. The current–current response function (a) and two approximations to the electron self-energy (b and c) are shown. The full and dashed lines represent electron and phonon Green’s functions, respectively. Self-consistent Green’s functions are used in a but not in b or c.

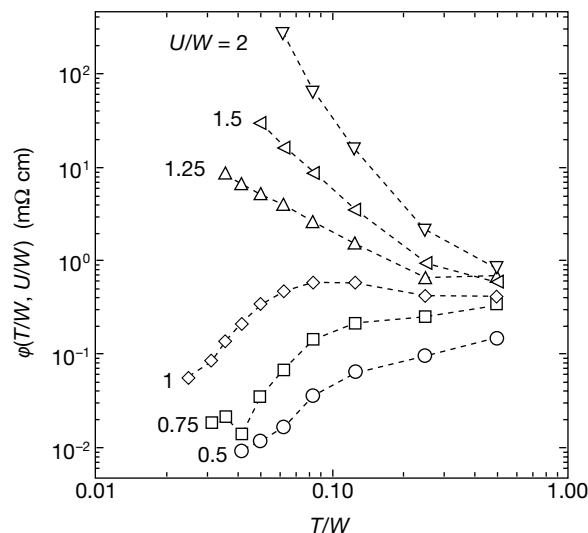


Figure 3 The resistivity of the non-degenerate Hubbard model. Results are shown for different values of the ratio between the Coulomb repulsion U and the band width W . The

figure illustrates how the resistivity saturates for large T in the electron–electron scattering model.

metallic case $\rho(T)$ saturates at $\rho \approx 0.4 \text{ m}\Omega \text{ cm}$, which corresponds to $l/d \approx 1/3$. Thus, in contrast to the electron–phonon scattering case, l is not very much smaller than d in the metallic state of this model and within DMFT.

To understand this, we study the electron self-energy Σ to second order in U , since Σ determines ρ in the DMFT. For low T , there is little scattering due to the small phase space available, as controlled by the Fermi functions. As T increases, the available phase space grows and ρ increases. However, for large T , ρ essentially saturates, because the Fermi functions approach a constant value. This is in strong contrast to the Bose occupation numbers (equation (3)), which keep increasing with T . The qualitative difference between the two models for large T can then be traced to the difference between fermions and bosons.

We now address the validity of the Boltzmann equation for the case of the electron–phonon scattering in view of $l \ll d$. We have calculated the resistivity using the Ziman version of the Bloch–Grüneisen solution of the Boltzmann equation¹⁸ and added the resistivity due to the orientational disorder as a T -independent contribution¹⁹. Using the plasma frequency $\omega_p = 1.2 \text{ eV}$, we obtain the dotted line in Fig. 1. The Boltzmann result is larger than the QMC result for large T , but there is no qualitative breakdown of the Boltzmann equation, although $l \ll d$ and the quasi-particle concept is not applicable. The justification for the Boltzmann equation in this limit is not the semi-classical derivation, but the (approximate) derivation from the full quantum mechanical Kubo formulation (equation (1)). The proper language is not in terms of a very short mean free path, but in terms of a very broad spectral function ($\text{Im } \Sigma$ large), as discussed above.

The QMC calculation gives an approximately linear T dependence. This agrees with the experimental result that ρ is linear down to about 100–200 K (ref. 8), although the linearity is not as pronounced as for some high-temperature superconductors. The result may seem surprising, as at small T the probability of exciting finite energy phonons is exponentially small, as is the contribution to ρ . Calculating the bubble diagram in Fig. 2a using electron Green's functions with the self-energy in Fig. 2b does indeed give an exponential behaviour. The use of the QMC Green's function, however, gives a linear behaviour, in agreement with the full QMC calculation. The reason is that the QMC Green's function also involves processes like those in Fig. 2c, where a virtual phonon is created followed by the decay of this phonon into an electron–hole pair²⁰. The excitation energy of such a pair can be arbitrarily small,

leading to a quadratic T -dependence for ρ (ref. 20). In our model, this quadratic behaviour goes over in an approximately linear behaviour already for very small T .

Calculations for different values of λ show a strong λ -dependence for small T . This is in agreement with the strong pressure dependence observed experimentally⁸, as the application of pressure increases the band width and reduces $N(0)$ and thereby λ . For $\lambda \approx 1$ there is a transition to an insulating state.

Our use of a fairly realistic model with a threefold degenerate electronic level and a fivefold degenerate phonon is not crucial for the qualitative behaviour of $\rho(T)$, such as $l \ll d$ for large T . An essential feature is, however, the use of intramolecular phonons acting on a narrow band. The phonon-induced shifts ($\sim T$) of the t_{1u} levels can be substantial compared with the small band width. It would be interesting to perform calculations for a model where the phonons couple to the hopping matrix elements and where inter-band transitions become important for large T . Such a model would be more relevant for transition metal compounds, for example, where the resistivity typically saturates when $l \approx d$. □

Received 17 November 1999; accepted 15 May 2000.

1. Fisk, Z. & Webb, W. B. Saturation of the high-temperature normal-state electrical resistivity of superconductors. *Phys. Rev. Lett.* **36**, 1084–1086 (1976).
2. Allen, P. B. in *Superconductivity in d- and f-Band Metals* (eds Suhl, H. & Maple, M. B.) 291–304 (Academic, New York, 1980).
3. Hebard, A. F., Palstra, T. T. M., Haddon, R. C. & Fleming, R. M. Absence of saturation in the normal-state resistivity of thin films of K_3C_{60} and Rb_3C_{60} . *Phys. Rev. B* **48**, 9945–9948 (1993).
4. Degiorgi, L. *et al.* Optical response of the superconducting state of K_3C_{60} and Rb_3C_{60} . *Phys. Rev. Lett.* **69**, 2987–2990 (1992).
5. Hou, J. G. *et al.* Determination of superconducting and normal state parameters of single crystal K_3C_{60} . *Solid State Commun.* **86**, 643–646 (1993).
6. Hou, J. G. *et al.* Resistivity saturation in alkali-doped C_{60} . *Solid State Commun.* **93**, 973–977 (1995).
7. Gunnarsson, O. Superconductivity of Fullerenes. *Rev. Mod. Phys.* **69**, 575–606 (1997).
8. Vareka, W. A. & Zettl, A. Linear temperature dependent resistivity in Rb_3C_{60} . *Phys. Rev. Lett.* **72**, 4121–4124 (1994).
9. Satpathy, S. *et al.* Conduction-band structure of alkali-doped C_{60} . *Phys. Rev. B* **46**, 1773–1792 (1992).
10. Mazin, I. I. *et al.* Orientational order in A_3C_{60} . Antiferromagnetic Ising model for the fcc lattice. *Phys. Rev. Lett.* **26**, 4142–4145 (1993).
11. Stephens, P. W. *et al.* Structure of single-phase superconducting K_3C_{60} . *Nature* **351**, 632–634 (1991).
12. Blankenbecler, R., Scalapino, D. J. & Sugar, R. L. Monte Carlo calculations of coupled boson-fermion systems. I. *Phys. Rev. D* **24**, 2278–2286 (1981).
13. Jarrell, M. & Gubernatis, J. E. Bayesian inference and the analytical continuation of imaginary-time quantum Monte Carlo data. *Phys. Rep.* **269**, 133–195 (1996).
14. Millis, A. J., Hu, J. & Das Sarma, S. Resistivity saturation revisited: Results from a dynamical mean field theory. *Phys. Rev. Lett.* **82**, 2354–2357 (1999).
15. Georges, A. *et al.* Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Rev. Mod. Phys.* **68**, 13–125 (1996).
16. Holstein, T. Theory of transport phenomena in an electron–phonon gas. *Ann. Phys.* **29**, 410–535 (1964).

17. Pálsson, G. & Kotliar, G. Thermoelectric response near the density driven Mott transition. *Phys. Rev. Lett.* **80**, 4775–4778 (1998).
18. Grimvall, G. *The Electron-Phonon Interaction in Metals* 212 (North-Holland, Amsterdam, 1981).
19. Gelfand, M. P. & Lu, J. P. Orientational disorder and normal-state electrical-transport properties of A_3C_{60} . *Phys. Rev. B* **46**, 4367–4370 (1992).
20. Allen, P. B. & Silbergliitt, R. Some effects of phonon dynamics on electron lifetime, mass renormalization, and superconducting transition temperature. *Phys. Rev. B* **9**, 4733–4741 (1974).

Acknowledgements

We are grateful to M. Jarrell for making his MaxEnt program available and E. Koch for a careful reading of the manuscript. The work has been supported by the Max-Planck-Forschungspreis.

Correspondence and requests for materials should be addressed to O.G. (e-mail: gunnar@and.mpi-stuttgart.mpg.de).

Collapse of stiff conjugated polymers with chemical defects into ordered, cylindrical conformations

Dehong Hu*, Ji Yu*, Kim Wong*, Biman Bagchi†, Peter J. Rossky* & Paul F. Barbara*

* Department of Chemistry & Biochemistry, University of Texas, Austin, Texas 78712, USA

† Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore 560012, India

The optical, electronic and mechanical properties of synthetic and biological materials consisting of polymer chains depend sensitively on the conformation adopted by these chains. The range of conformations available to such systems has accordingly been of intense fundamental^{1,2} as well as practical^{3–6} interest, and distinct conformational classes have been predicted, depending on the stiffness of the polymer chains and the strength of attractive interactions between segments within a chain^{7–10}. For example, flexible polymers should adopt highly disordered conformations

resembling either a random coil or, in the presence of strong intrachain attractions, a so-called ‘molten globule’^{2,10}. Stiff polymers with strong intrachain interactions, in contrast, are expected to collapse into conformations with long-range order, in the shape of toroids or rod-like structures^{8,9,11}. Here we use computer simulations to show that the anisotropy distribution obtained from polarization spectroscopy measurements on individual poly[2-methoxy-5-(2'-ethylhexyl)oxy-1,4-phenylenevinylene] polymer molecules is consistent with this prototypical stiff conjugated polymer adopting a highly ordered, collapsed conformation that cannot be correlated with ideal toroid or rod structures. We find that the presence of so-called ‘tetrahedral chemical defects’, where conjugated carbon–carbon links are replaced by tetrahedral links, divides the polymer chain into structurally identifiable quasi-straight segments that allow the molecule to adopt cylindrical conformations. Indeed, highly ordered, cylindrical conformations may be a critical factor in dictating the extraordinary photophysical properties of conjugated polymers, including highly efficient intramolecular energy transfer and significant local optical anisotropy in thin films.

Figure 1 illustrates a range of possible conformations that can be adopted by polymer chains. To investigate the conformational properties of poly[2-methoxy-5-(2'-ethylhexyl)oxy-1,4-phenylenevinylene] (MEH-PPV), a widely studied prototype of stiff conjugated polymers, we use a single-molecule spectroscopy approach^{12–15} and MEH-PPV samples that have an average molecular weight of 453,000 with 1,700 repeat units in a typical polymer molecule¹⁶. As a result of conformational distortions^{17,18}, chemical defects along the polymer chain¹⁹, and electron correlation effects²⁰, the lowest-energy optical absorption of an isolated MEH-PPV molecule is due to about 140 different local ‘quasi-chromophores’ along the chain, each with a length of 10–17 repeat units²¹. For each chromophore, the π – π^* absorption is polarized along the direction of the local segment of the chain²². Although there is a distribution of repeat-unit lengths and spectroscopic transition energies, the optical transitions are overlapped near the absorption peak (~ 500 nm) due to vibronic broadening²³. There is convincing evidence that the majority of the emission actually occurs from a small region on the polymer chain, due to rapid and efficient intramolecular electronic energy transfer^{16,19}.



Figure 1 Typical conformations of a 100-segment homopolymer generated by Monte Carlo simulations. The simulation parameters are summarized in Table 1. They are

denoted as: **I**, random coil; **II**, molten globule; **III**, toroid; **IV**, rod; **V**, defect-coil; and **VI**, defect-cylinder.

The MEH-PPV molecules are isolated at low concentration in 300-nm-thick, spin-coated inert polymer film (typically polycarbonate) on a glass cover-slip substrate. The MEH-PPV molecules are laterally separated in the film plane (denoted by x - y) of the sample on average by 3 μm . Individual molecules are optically excited with focused (spot size ~ 400 nm), linearly polarized light propagating in the z direction using an oil immersion objective located on the substrate side of the sample. The MEH-PPV emission is collected along the z direction with the same objective used for excitation, filtered to remove scattered excitation light, and detected with an avalanche photodiode detector, all with a minimal bias for polarization. The direction of polarization, θ , of the excitation light is linearly temporally modulated between 0 and π with a period of one second by an electro-optic modulator²⁴. The total unpolarized fluorescence intensity, $I(\theta)$, is synchronously recorded for 30 cycles of the modulation, corresponding to 10^7 excitations of MEH-PPV.

$I(\theta)$ is observed to be invariant with cycle within experimental error during this averaging process for the vast majority of the MEH-PPV molecules. Time- and wavelength-resolved spectroscopy of individual polymer molecules demonstrates that $I(\theta)$ is not distorted by spectral diffusion, blinking or other photo-induced processes. The invariance of $I(\theta)$ implies that the characteristic chain conformation of each molecule is fixed on the experimental timescale. A typical averaged $I(\theta)$ is shown as the inset in Fig. 2. It is important to emphasize that $I(\theta)$ reports the polarization dependence of the excitation (absorption) process through the emission intensity. The quasi-chromophore can be approximated by a transition dipole along the local direction of the polymer chain²², so that the absorption is due to a set of about 140 comparable transition dipoles, each oriented along a particular direction in the x - y - z laboratory coordinate system. The corresponding absorption properties of a single polymer molecule can be represented by an "absorption ellipsoid" in the molecular frame (x' , y' , z') with three distinct orthogonal absorption cross-sections²⁵: C_x , C_y , and C_z . $I(\theta)$ is the projection of this absorption ellipsoid in the x - y plane, as the propagation direction is along z . The functional form of $I(\theta)$ is thus

$$I(\theta) \propto 1 + M \cos 2(\theta - \phi) \quad (1)$$

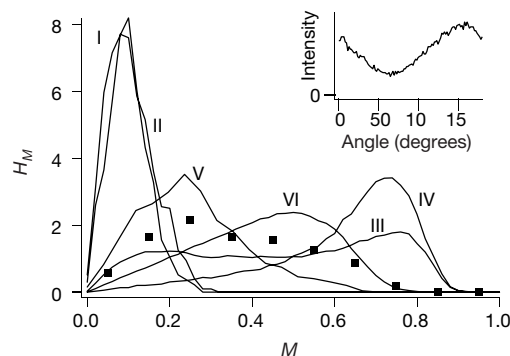


Figure 2 The distribution H_M of modulation depths M from single-molecule polarization spectroscopy and from Monte Carlo simulations. The squares show the distribution for 102 MEH-PPV single molecules measured from polarized fluorescence excitation intensity. The lines are the simulated distributions for the six polymer conformational types, shown in Fig. 1. The simulation parameters are given in Table 1. Inset, $I(\theta)$ for a single MEH-PPV molecule. The film was prepared by spin coating from a toluene solution. Oxygen was excluded from the film by outgassing under vacuum and coating with a 200-nm Al layer. Samples (ambient temperature) were excited at 488 nm using a numerical aperture of 0.7 and a laser power of 50 nW. Emission was collected with a numerical aperture of 1.25.

where the modulation depth M is the anisotropy of the absorption ellipsoid projected on the x - y plane, and ϕ is the orientation of maximum absorption. Our experimental $I(\theta)$ (inset in Fig. 2) is well fitted by equation (1). The observed ϕ values are randomly distributed between 0 and π , as expected.

To determine the absorption ellipsoid for a specific polymer chain, it would be necessary to make measurements with light propagation along multiple axes. However, this is not possible with our apparatus. We rely instead on the distribution of molecular orientations of the entire polymer molecule to achieve this goal. We record M for a large number of single molecules and construct a histogram of M values (H_M as shown in Fig. 2.) The ensemble of values of the single polymer molecule property M reflects the distribution of chain orientations, chain conformations, and molecular weights. The physical significance of H_M can be established by considering limiting cases. For a straight polymer chain, the individual transition dipoles fall on a straight line, and H_M is a delta function located at $M = 1$. In the opposite limit, that is, a perfectly isotropic distribution of dipoles (for example, a random coil of infinite chain length) the absorption ellipsoid is a sphere, and H_M is a delta function located at $M = 0$. We have experimentally verified the latter case by studying approximately 10^4 randomly oriented dye molecules embedded in a 100-nm latex sphere. Those data demonstrate that there is an experimental error of about 0.1 in the M axis of Fig. 2.

We have used Monte Carlo simulation of beads on a chain (bond fluctuation method^{26,27}) to generate an ensemble of conformations for model polymer chains with different numbers of chain defects, intersegment attraction, and chain stiffness. An individual bead or segment represents a small portion of the polymer chain, that is, 2.5 repeat units, which corresponds to 1.5 nm. Typically, a chain with 100 beads was used in the simulations representing a MEH-PPV molecule with 250 repeat units. (Simulations with more beads were prohibitively time-consuming.) The intrachain interactions were modelled with a Lennard-Jones-like attractive potential between a pair of beads with depth E_{cc} . Simulations with a square-well intersegment potential gave analogous results. The chain bending energy E_{bend} is $b\alpha^2$, where α is the bending angle. The equilibrium bond angle between adjacent bonds was assumed to be 180° except in the case of defects, which are discussed below. The bending force constant, b , was estimated from published force fields for solid polymer²⁸. We note that the intrachain interaction model gives a differential free energy for the polymer in a solvent, arising in part

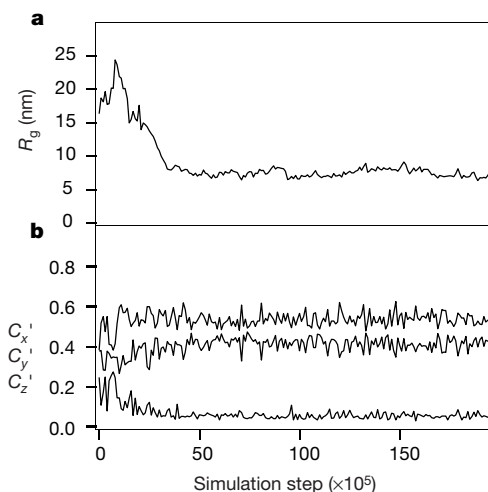


Figure 3 The structural collapse trajectory of a random coil to a toroid. **a**, The radius of gyration R_g ; and **b**, the normalized absorption cross-sections along the three principal axes of the conformation are plotted versus Monte Carlo simulation steps.

Table 1 Conformations and simulation parameters of homopolymers

Symbol	Type	E_{cc}^* (kT)	b (kT rad ⁻²)	Number of defects	$\langle M^2 \rangle^{1/2} \dagger$	$\langle R_g^2 \rangle^{1/2} \ddagger$ (nm)
I	Random coil	0	0	N/A	0.11	13
II	Molten globule	0.6	0	N/A	0.12	5.0
III	Toroid	0.6	10	0	0.53	7.4
IV	Rod§	0.6	10	0	0.67	10
V	Defect-coil	0	10	15	0.28	21
VI	Defect-cylinder	0.6	10	15	0.46	7.2

All simulations were conducted on the same cubic three-dimensional lattice. The lattice constant is 0.3 nm. The linked bead distance is restricted from 1.2 nm to 1.8 nm (4 to 6 lattice units). The bead displacement is one lattice unit. The simulations are usually conducted for 10^8 Monte Carlo steps. A fixed number of defects are at random locations of the chain. Harmonic bending potential: $E_{\text{bend}} = b\alpha^2$. N/A, not applicable.

* E_{cc} is the depth of a Lennard-Jones attraction potential between bead pairs of the polymer chain. The bead attraction potential is:

$$\begin{cases} E_{LJ} = \infty & \text{if } l \leq 2.1 \text{ nm} \\ E_{LJ} = 4E_{cc} \left(\left(\frac{2.1}{l} \right)^{12} - \left(\frac{2.1}{l} \right)^6 \right) & \text{if } 2.1 \text{ nm} < l \leq 4.2 \text{ nm} \\ E_{LJ} = 0 & \text{if } l > 4.2 \text{ nm} \end{cases}$$

$\dagger \langle M^2 \rangle^{1/2}$ is the root mean square of the modulation depth of anisotropy.

$\ddagger \langle R_g^2 \rangle^{1/2}$ is the root mean square of radius of gyration.

§ Rods composed of two to six straight-chain regions are observed in the simulations. R_g strongly depends upon the number of straight-chain regions in a rod and the distribution among types of rods is not converged in our limited number of simulations (20 trials). To avoid this source of error in Table 1, we restrict our analysis of R_g and M to rods with four straight-chain regions.

|| This number corresponds to 15 defects per polymer chain of 100 beads. Since the stiffness parameter (b) has been chosen such that each bead represents 2.5 repeat units, a polymer chain with 15 defects actually corresponds to 6% of the underlying 250 units.

from the difference between interactions among polymer segments compared to polymer with solvent. In general, this free energy will be dependent on solute conformation. For a polymeric solvent, as exists in the present films, one might expect this dependence to be more significant than for a simple solvent, due to the conformational coupling of solvent and solute. This additional complexity is not captured in the highly simplified polymer models used in the present investigation.

For each parameter set, an ensemble of 20 chain conformations was generated. Each conformation is from a distinct initial conformation after about 10^8 Monte Carlo steps (one step is one attempted move for each bead of the chain). The set of conformations were used to calculate the ensemble averaged: radius of gyration, R_g ; laboratory frame anisotropy (or modulation depth), M (Table 1); and H_M . For the anisotropy calculations, it was assumed that each local transition dipole was along the bond connecting the adjacent beads.

The simplest example we consider is the flexible chain ($b = 0$) with no attraction ($E_{cc} = 0$), which adopts the well known self-avoiding random-coil conformation. A typical example of a random-coil conformation from the simulation is shown in Fig. 1, structure I. The random coil conformation has an H_M distribution (curve I in Fig. 2) that is considerably less anisotropic than the experimental results. The average anisotropy of 0.11 (Table 1) is consistent with theory for the random coil, predicting that $\langle M^2 \rangle^{1/2} \propto N^{1/2}$ where $N+1$ is the number of beads²⁹. It is well established^{1,2} that chain-chain attraction of sufficient magnitude will induce a flexible chain to collapse to a compact conformation. A typical molten-globule conformation from the simulations ($E_{cc} = 0.6$ kT, $b = 0$) is shown by structure II in Fig. 1. The H_M curve for the molten globule is similar to that for the random coil, reflecting the “liquid-like” disorder in the globule¹⁰, but the radius of gyration is smaller for the molten globule (Table 1).

The inclusion of chain stiffness has a dramatic effect on the structural and spectroscopic anisotropy. Simulations with $E_{cc} = 0.6$ kT, $b = 10$ kT rad⁻² reveal two highly ordered but distinct conformations: toroid and rod (Fig. 1)^{8,9,11}. They can be quantitatively identified by the absorption ellipsoid. The simulated H_M curves (Fig. 2) are more similar to the experimental results than are the simulated H_M curves for the random coil and molten globule. The results for the toroid and rod are, however, peaked toward M values higher than experiment and tend to have a narrower distribution. Further insight into the rod and toroid conformations can be obtained by considering a representative simulation trajectory (Fig. 3). For this trajectory, a random-coil

($E_{cc} = 0$, $b = 0$) conformation was chosen as an initial conformation of the simulation, and the parameters were switched to values for a stiff chain with attraction ($E_{cc} = 0.6$ kT, $b = 10$ kT rad⁻²) at the beginning of the simulation. After each 10^5 steps of the simulation, the instantaneous radius of gyration and absorption cross-sections C_x , C_y and C_z (molecular frame) were calculated. Near the beginning of the simulation, the conformations elongated in response to the increase in chain stiffness, as evidenced by a large R_g . After further Monte Carlo steps, thermal fluctuation produced segment-segment contacts that ‘seeded’ a chain collapse as reflected by the drop in R_g . The instantaneous absorption cross-sections easily allow for the assignment of the chain conformation. Initially, the near equality of the three cross-sections is consistent with the random-coil conformation. Chain collapse to a toroid-like structure is evidenced by one small and two similar, but larger, cross-sections.

One important structural feature that has been left out of existing simulation and theory of stiff polymer chains is chemical defects along the chain. The synthetic procedure for MEH-PPV chains results in chains with 1% or greater tetrahedral rather than conjugated links. Conjugated polymers tend to be reactive, which allows for the further introduction of tetrahedral defects during photoexcitation. In fact, spectroscopic analysis of our MEH-PPV sample leads to a defect concentration estimate of ~5% (ref. 19). We have introduced such chemical defects at random beads in the simulation by tetrahedral bond angles (unchanged force constant). The inclusion of defects has a profound effect on the conformations and their anisotropies. For the case of a stiff chain with defects but no intersegment attraction, an extended conformation is found (V, Fig. 1) with approximately straight chain segments between defects, denoted here as a ‘defect-coil’. The equilibrium structure (not shown) of the same chain without defects is highly extended and considerably more anisotropic. The defect-coil has a lower average anisotropy than the experimental behaviour of MEH-PPV and peaked at lower values.

Structure VI in Fig. 1 is from simulations of stiff chains with defects and intersegment attraction. The simulated anisotropy for this roughly cylindrical conformation is more similar to experiment than any other conformation of Fig. 1. Structure VI, which we denote as a defect-cylinder, is arguably the most accurate representation of the realistic collapsed conformations of conjugated polymers such as MEH-PPV among the idealized structures that have been considered. Indeed, it is unlikely that real conjugated polymer chains, except in special cases, would be sufficiently free of defects to fold into the toroid or rod structures. With respect to the

experimental samples in this study, separate spectroscopic experiments strongly suggest that the spin-coating conditions we use to prepare the samples produce films that contain roughly equal amounts of collapsed and non-collapsed conformations. In fact, an equal superposition of the anisotropy distributions H_M of defect-coil and defect-cylinder (V and VI) is in reasonable agreement with the experimental data. Interestingly, the discovery of highly ordered, cylindrical conformations reported here for MEH-PPV may be a critical factor resolving the puzzling properties of MEH-PPV and related polymers. These include the significant local anisotropy of thin films of pure materials²⁴, and evidence that the structure of MEH-PPV in films can be controlled by the conformation in the solution used to spin-coat the films³⁰. The new conformations introduced in the present work, the defect-coil (V) and defect-cylinder (VI), exhibit a key characteristic, namely, structurally identifiable quasi-straight-chain segments. These segments are therefore excellent candidates for the localized quasi-chromophores in conjugated polymers. These conformations probably also describe other stiff-chain conjugated systems, such as polyfluorene and poly(phenyleneethynylene), that can develop the necessary tetrahedral defects. □

Received 10 February; accepted 5 May 2000.

- de Gennes, P.-G. *Scaling Concepts In Polymer Physics* (Cornell Univ. Press, Ithaca, New York, 1979).
- Grosberg, A. Y. & Kuznetsov, D. V. Quantitative theory of the globule-to-coil transition. *Macromolecules* **25**, 1970–2003 (1992).
- Friend, R. H. *et al.* Electroluminescence in conjugated polymers. *Nature* **397**, 121–128 (1999).
- Hide, F., Diazgarcia, M. A., Schwartz, B. J. & Heeger, A. J. New developments in the photonic applications of conjugated polymers. *Acc. Chem. Res.* **30**, 430–436 (1997).
- Yang, C. Y., Hide, F., Diazgarcia, M. A., Heeger, A. J. & Cao, Y. Microstructure of thin films of photoluminescent semiconducting polymers. *Polymer* **39**, 2299–2304 (1998).
- Bloomfield, V. A. Condensation of DNA by multivalent cations: Consideration on mechanism. *Biopolymers* **31**, 1471–1481 (1991).
- Grosberg, A. Y. Certain possible conformational states of a uniform elastic polymer chain. *Biophysics* **24**, 30–36 (1979).
- Ivanov, V. A., Paul, W. & Binder, K. Finite chain length effects on the coil-globule transition of stiff-chain macromolecules—a Monte Carlo simulation. *J. Chem. Phys.* **109**, 5659–5669 (1998).
- Noguchi, H. & Yoshikawa, K. Morphological variation in a collapsed single homopolymer chain. *J. Chem. Phys.* **109**, 5070–5077 (1998).
- Zhou, Y. Q., Karplus, M., Wichter, J. M. & Hall, C. K. Equilibrium thermodynamics of homopolymers and clusters - molecular dynamics and Monte Carlo simulations of systems with square-well interactions. *J. Chem. Phys.* **107**, 10691–10708 (1997).
- Kuznetsov, Y. A. & Timoshenko, E. G. On the conformational structure of a stiff homopolymer. *J. Chem. Phys.* **111**, 3744–3752 (1999).
- Xie, X. S. & Trautman, J. K. Optical studies of single molecules at room temperature. *Annu. Rev. Phys. Chem.* **49**, 441–480 (1998).
- Basche, T., Moerner, W. E., Orrit, M. & Wild, U. P. (eds) *Single Molecule Optical Detection, Imaging, and Spectroscopy* (Verlag Chemie, Munich, 1996).
- Ha, T., Laurence, T. A., Chmela, D. S. & Weiss, S. Polarization spectroscopy of single fluorescent molecules. *J. Phys. Chem. B* **103**, 6839–6850 (1999).
- Vandenbout, D. A. *et al.* Discrete intensity jumps and intramolecular electronic energy transfer in the spectroscopy of single conjugated polymer molecules. *Science* **277**, 1074–1077 (1997).
- Hu, D., Yu, J. & Barbara, P. F. Single-molecule spectroscopy of the conjugated polymer MEH-PPV. *J. Am. Chem. Soc.* **121**, 6936–6937 (1999).
- Yaliraki, S. N. & Silbey, R. J. Conformational disorder of conjugated polymers—implications for optical properties. *J. Chem. Phys.* **104**, 1245–1253 (1996).
- Gettinger, C. L., Heeger, A. J., Drake, J. M. & Pine, D. J. A photoluminescence study of poly(phenylene vinylene) derivatives—the effect of intrinsic persistence length. *J. Chem. Phys.* **101**, 1673–1678 (1994).
- Padmanaban, G. & Ramakrishnan, S. Conjugation length control in soluble poly[2-methoxy-5-(2'-ethylhexyl)oxy-1,4-phenylenevinylene] (MEHPPV): synthesis, optical properties, and energy migration. *J. Am. Chem. Soc.* **122**, 2244–2251 (2000).
- Mukamel, S., Tretiak, S., Wagersreiter, T. & Chernyak, V. Electronic coherence and collective optical excitations of conjugated molecules. *Science* **277**, 781–787 (1997).
- Woo, H. S. *et al.* Optical spectra and excitations in phenylene vinylene oligomers. *Synth. Met.* **59**, 13–28 (1993).
- Hagler, T. W., Pakbaz, K. & Heeger, A. J. Polarized-electroabsorption spectroscopy of a soluble derivative of poly(p-phenylenevinylene) oriented by gel processing in polyethylene—polarization anisotropy, the off-axis dipole moment, and excited-state delocalization. *Phys. Rev. B* **49**, 10968–10975 (1994).
- Bassler, H. & Schweitzer, B. Site-selective fluorescence spectroscopy of conjugated polymer and oligomers. *Acc. Chem. Res.* **32**, 173–182 (1999).
- Blatchford, J. W. *et al.* Spatially and temporally resolved emission from aggregates in conjugated polymers. *Phys. Rev. B* **54**, R3683–R3686 (1996).
- Thulstrup, E. W. & Michl, J. *Elementary Polarization Spectroscopy* (VCH, New York, 1989).
- Carmesin, I. & Kremer, K. The bond fluctuation method: a new effective algorithm for the dynamics of polymers in all spatial dimensions. *Macromolecules* **21**, 2819–2823 (1988).
- Helfand, E. Theory of the kinetics of conformational transitions in polymer. *J. Chem. Phys.* **54**, 4651–4661 (1971).

- Orion, I., Buisson, J. P. & Lefrant, S. Spectroscopic studies of polaronic and bipolaronic species in n-doped poly(paraphenylenevinylene). *Phys. Rev. B* **57**, 7050–7065 (1998).
- Lodge, T. P. & Fredrickson, G. H. Optical anisotropy of tethered chains. *Macromolecules* **25**, 5643–5650 (1992).
- Nguyen, T. Q., Doan, V. & Schwartz, B. J. Conjugated polymer aggregates in solution: Control of interchain interactions. *J. Chem. Phys.* **110**, 4068–4078 (1999).

Acknowledgements

This work was supported by grants from the National Science Foundation (P.F.B.), the Robert A. Welch Foundation (P.J.R. & P.F.B.), and the Texas Advanced Research Program (P.J.R.). Further support was provided by the Institute for Theoretical Chemistry and by the Laboratory for Spectroscopic Imaging, University of Texas. We also thank A. Yethiraj for discussions.

Correspondence and requests for materials should be addressed to P.F.B. (e-mail: p.barbara@mail.utexas.edu).

Dynamic self-assembly of magnetized, millimetre-sized objects rotating at a liquid–air interface

Bartosz A. Grzybowski*, Howard A. Stone† & George M. Whitesides*

* Harvard University, Department of Chemistry and Chemical Biology, 12 Oxford Street, Cambridge, Massachusetts 02138, USA

† Harvard University, Division of Engineering and Applied Sciences, Pierce Hall, Cambridge, Massachusetts 02138, USA

Spontaneous pattern formation by self-assembly is of long-standing^{1–3} and continuing interest^{4,5} not only for its aesthetic appeal^{6,7}, but also for its fundamental^{8–18} and technological relevance¹⁹. So far, the study of self-organization processes has mainly focused on static structures, but dynamic systems^{20–22}—those that develop order only when dissipating energy—are of particular interest for studying complex behaviour^{23,24}. Here we describe the formation of dynamic patterns of millimetre-sized magnetic disks at a liquid–air interface, subject to a magnetic field produced by a rotating permanent magnet. The disks spin around their axes with angular frequency equal to that of the magnet, and are attracted towards its axis of rotation while repelling each other. This repulsive hydrodynamic interaction is due to fluid motion associated with spinning; the interplay between attractive and repulsive interactions leads to the formation of patterns exhibiting various types of ordering, some of which are entirely new. This versatile system should lead to a better understanding of dynamic self-assembly, while providing a test-bed for stability theories of interacting point vortices^{25–28} and vortex patches²⁹.

We fabricated circular disks by filling hollow polyethylene tubing (~1 mm inside diameter, ~2 mm outside diameter) with poly(dimethylsiloxane) (PDMS) doped with magnetite (~5–30 wt%), and cutting the resulting composite into slices ~400 μ m thick. A permanent bar magnet of approximate dimensions ($L \times W \times D$, in cm) $5.6 \times 4 \times 1$ was placed at a distance h (about 2–4 cm) below the interface, and rotated with angular velocity ω (Fig. 1a). The magnet was magnetized along its longest dimension, and had magnetization $M \approx 1,000 \text{ G cm}^{-3}$. When the magnet was stationary, the disks were attracted towards its poles, where they formed orderless aggregates. When the magnet rotated, the disks were drawn towards its axis of rotation. In addition, the magnetic moments of the disks interacted with the rotating magnetic field, and the disks spun around their centres. The fluid motion associated with spinning resulted in a repulsive hydrodynamic

interaction between the disks.

The competition between axisymmetric magnetic attraction and hydrodynamic repulsion of the disks led to the formation of ordered aggregates (Fig. 2). We denote the number of disks in an aggregate by n . One disk ($n = 1$) experiences only the attractive magnetic force, and in its stable position, its centre coincides with the axis of rotation of the magnet (we will use “the centre” to mean this axis of rotation). For $n = 2$, the disks repel each other, and are equidistant from the centre, around which they slowly precess (angular velocity $\Omega \approx 2$ r.p.m.). For $n = 3-5$, the disks organize into polygons (a triangle for $n = 3$, a square for $n = 4$, and a pentagon for $n = 5$) precessing around the centre. For $n = 6$, one disk goes to the centre, while the remaining five form a pentagon around it. For $n > 6$, multishell structures appear. The shells precess with the same angular velocity Ω . Most of the aggregates have one, well defined stable structure, but in three cases we observed polymorphs, where multiple steady states are possible. For 10 and 12 disks, the patterns spontaneously interconvert between the polymorphs. In the 19-membered aggregate, the hexagonal structure (shown on

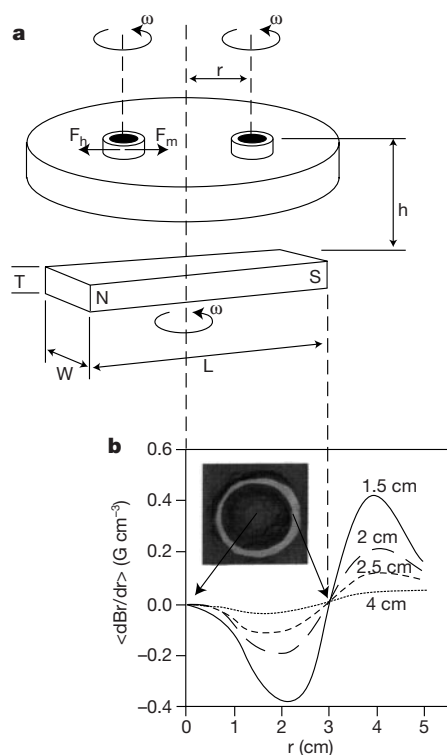


Figure 1 Experimental set-up and magnetic force profiles. **a**, A scheme of the experimental set-up. A bar magnet rotates at angular velocity ω below a dish filled with liquid (typically ethylene glycol/water or glycerine/water solutions). Magnetically doped disks are placed on the liquid–air interface, and are fully immersed in the liquid except for their top surface. The disks spin at angular velocity ω around their axes. A magnetic force F_m attracts the disks towards the centre of the dish, and a hydrodynamic force F_h pushes them apart from each other. The curves in **b** give the time average (over one revolution of the magnet) of the radial derivative of the magnetic induction $\partial B_r/\partial r$ as a function of r . This derivative is proportional to the radially directed magnetic force $F_m(r)$ acting on a ferromagnetic point object located at r . Each curve corresponds to a different separation h between the upper face of the magnet and the liquid–air interface. The force is ‘attractive’ (directed towards the centre) for $r < \sim 3$ cm (roughly equal to $L/2$, half of the length L of the magnet), and ‘repulsive’ otherwise. The patterns are stable only within the attractive region. Inset, an optical micrograph showing the results of an experiment with iron filings suspended in a viscous oil, and subjected to a rotating magnetic field such as that used in the calculations: the annular band where F_m changes sign (near $r \approx L/2$) is depleted of filings.

the left in Fig. 2) appears only above a threshold rotational speed $\omega \approx 800$ r.p.m., while the less symmetric structure (shown on the right) exists below this value (although the hexagonal structure can be sustained in a metastable equilibrium at $\omega < 800$ r.p.m. by slowly decreasing ω from above the threshold). The aggregate formed by 19 1.27-mm disks is the largest stable structure observed with the magnet used in our experiments. The reason for this limitation is

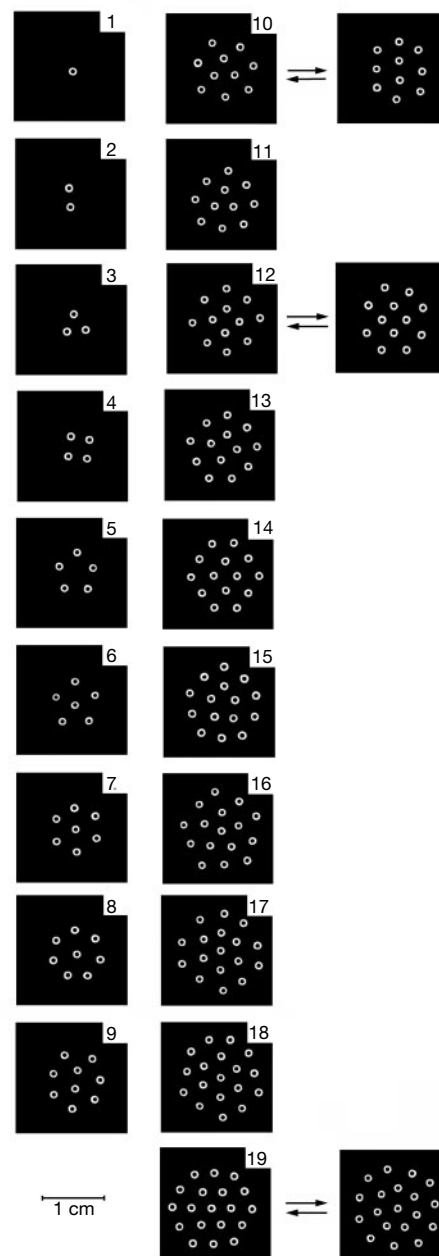


Figure 2 Dynamic patterns formed by various numbers (n) of disks rotating at the ethylene glycol/water–air interface. This interface is 27 mm above the plane of the external magnet. The disks are composed of a section of polyethylene tube (white) of outer diameter 1.27 mm, filled with poly(dimethylsiloxane), PDMS, doped with 25 wt% of magnetite (black centre). All disks spin around their centres at $\omega = 700$ r.p.m., and the entire aggregate slowly ($\Omega < 2$ r.p.m.) precesses around its centre. For $n < 5$, the aggregates do not have a ‘nucleus’—all disks are precessing on the rim of a circle. For $n > 5$, nucleated structures appear. For $n = 10$ and $n = 12$, the patterns are bistable in the sense that the two observed patterns interconvert irregularly with time. For $n = 19$, the hexagonal pattern (left) appears only above $\omega \approx 800$ r.p.m., but can be ‘annealed’ down to 700 r.p.m. by slowly decreasing the spinning rate. Without annealing, a less symmetric pattern exists at $\omega = 700$ r.p.m.

that, as the aggregate gets bigger, the outermost disks experience a less-homogenous and weaker magnetic field than those closer to the centre. Under such conditions, they often fall out of resonance with the external rotating field and stop spinning.

The spacing between the nearest neighbours in an aggregate of size n is approximately independent of n , but increases with increasing ω . The faster a disk spins, the larger the hydrodynamic repulsion it exerts on other disks. Thus, for a given strength of the magnetic field, the separation between the disks increases with ω . Figure 3a shows two aggregates of 860- μm -diameter disks: in the picture on the left ($\omega \approx 800$ r.p.m.), the lattice spacing is smaller than in the picture on the right, where the disks are rotating at $\omega \approx 1,100$ r.p.m. (notice the outer disordered rim, where the disks are off-resonance with the magnetic field). The lattice spacing can be increased only to about seven disk diameters: above this separation, disks start moving independently from each other, and the ordering disappears (Fig. 3b).

The largest aggregates obtained in our experiments were hexagonal structures, shown in Fig. 3c. Here the disks were 570 μm in diameter; because of their small size, many disks could be packed into the region of homogeneous magnetic field. The magnetite content in these disks was low (~ 5 wt%), to minimize their tilting or tumbling. The structures were stable only when the packing in the aggregate was hexagonal, and the structure had a six-fold symmetry.

To understand better the nature and the capabilities of self-

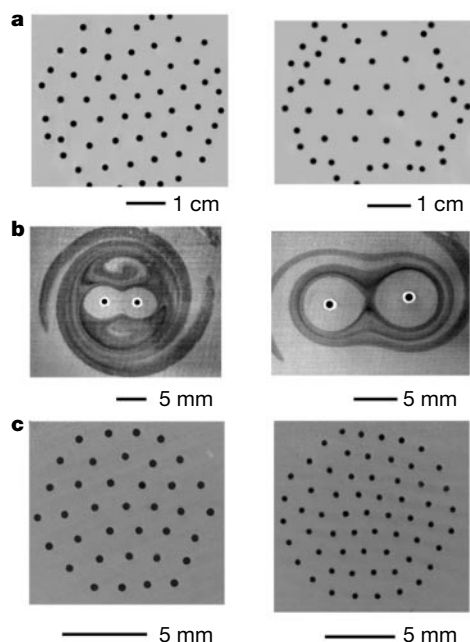


Figure 3 Illustrations of various effects controlling the dimensions and the stability of patterns. In **a**, the lattices were formed by 0.86-mm disks; the lattice spacing increased with the rotational speed (800 r.p.m. in the picture on the left, 1,100 r.p.m. in the picture on the right). The pictures in **b** illustrate the effect of ω on the stability of aggregates. Two 1.27-mm disks were spinning on the ethylene glycol–water interface. The streamlines were visualized by placing drops of rhodamine/water solution onto the interface. In the picture on the left, the disks were rotating at $\omega = 700$ r.p.m. No dye entered a high-pressure ‘8-shaped’ region connecting the rotating disks, and the separation between the disks did not change with time. When ω was increased to 1,100 r.p.m., the high-pressure regions produced by the disks became disjoint, as indicated by the crossing of the streamlines in the midpoint between the disks (right). Disks moved independently of each other, and the separation between them varied with time. Optical micrographs in **c** show hexagonally ordered aggregates formed by 570- μm PDMS disks doped with 5% magnetite, and rotating at $\omega = 1,100$ r.p.m. on a liquid–air interface 2.5 cm above the top face of the magnet. In this experiment, the liquid was a solution of 75% ethylene glycol:25% water.

assembly in our system, we investigated the forces involved in the process both theoretically and experimentally. We calculated the magnetic force acting on the disks using a standard current-sheet method (see Supplementary Information); profiles of the magnetic force are shown in Fig. 1b. The hydrodynamic repulsive forces arise from the combination of the spinning of a disk and its translation through the liquid. Consider the simplest aggregate composed of two disks. Each disk produces a rotational flow of liquid around it, and also experiences a flow produced by the other disk. Thus, each disk can be represented as rotating around its axis, while simultaneously translating through the flow with a gradient of shear produced by the other disk. For the typical rotation speeds (ω), disk radii (a), densities (ρ) and fluid shear viscosities (μ) of our experiments, the Reynolds number on the scale of the particle is $\text{Re} = (\rho\omega a^2/\mu) < O(1)$, although it is not very much smaller (here, we use the symbol $O(\epsilon)$ to mean, as usual, ‘order-of-magnitude’ of ϵ). We note that an explanation based on the Bernoulli principle appears inappropriate, because the Reynolds number is small.

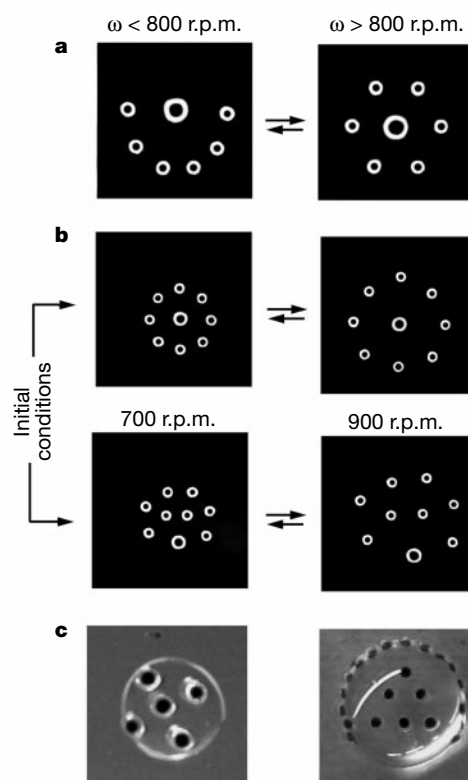


Figure 4 Various phenomena observed in systems of magnetized rotating disks. In **a** the two structures interconvert depending on the rotational speed ω of the external magnet. For $\omega < \sim 800$ r.p.m., both the small disks (1.27 mm in diameter), and the large disk (2.42 mm in diameter), rotate at ω ; the small disks organize in a semicircle slowly precessing around the large disk. When ω is increased, the large disk falls off resonance from the external magnetic field, and starts rotating at $\omega' < \omega$. As a result, the hydrodynamic repulsion it exerts on its smaller neighbours decreases; the structure of the aggregate changes, and resembles the one formed by seven equally sized disks (Fig. 2). In **b** the difference in size between the large disk (diameter 1.7 mm) and the small disks (1.27 mm) is not large enough to ensure that the large disk will always organize smaller disks around it; the type of aggregate formed depends on initial conditions. Once either of the two types is formed, its size can be modified by adjusting ω , but the two morphologies cannot be interconverted. The optical micrographs in **c** show aggregates assembled on curved surfaces; we did not observe such structures in our experiments on planar interfaces. The picture on the left shows a nucleated five-membered aggregate of 1.27-mm disks rotating on the top surface of a droplet of perfluorodecalin (PFD) covered with water. The picture on the right shows a pattern formed on the bottom of a water droplet floating on a PFD–air interface.

We thus expect the explanation for the repulsive hydrodynamic interaction to lie in the realm of viscosity-dominated flows, where small inertial effects are probably significant. There are only a few analytical results available for a force-free particle in a wall-bounded shear flow³⁰: such a particle experiences a lift force transverse to the streamlines $O(\mu u_C a \text{Re}_G)$, where u_C is the typical velocity on the scale of the particle, and $\text{Re}_G = \rho G a^2 / \mu$ is the Reynolds number based on the local shear rate G . For our system, the particle experiences a torque, which generates a local velocity $u_C = \omega a$, and the background shear rate is $G = O(\omega a^3 / d^3)$, where d is the distance between the centres of the disks. Therefore, the repulsive hydrodynamic force F_h between the disks is predicted to be $F_h = O(\mu(\omega a) a \text{Re}_G) = O(\rho \omega^2 a^7 / d^3)$. In the stable configuration, this hydrodynamic repulsion balances the magnetic force attracting the disks towards the axis of rotation of the external magnet. Although a detailed calculation remains to be done, we verified experimentally that qualitative trends predicted by this analysis appear to be correct (see Supplementary Information).

Our qualitative theory explains the nature of the pairwise forces acting in the system, but it is clearly insufficient to account for the emergence of the patterns that we observe, especially in more complicated systems. The structure of the aggregates of disks of different sizes can be controlled by adjusting the rotational speed to decouple larger disks from the field selectively (Fig. 4a). Aggregates that have structures dependent on the initial distribution of the disks can also be prepared (Fig. 4b). The self-assembly can take place on both planar and curved surfaces (Fig. 4c). It seems likely that changing the shape of the spinners, and engineering the profile of the magnetic field, would lead to even more complex behaviour.

We believe that this system could be a useful experimental tool in assisting theoretical research on multivortex flows. Also, if the size of the disks could be reduced to several micrometres, the type of self-assembly that we describe here might have practical applications in optics; for example, in tunable diffraction gratings or in photonic bandgap materials. If it was found to be possible to 'solidify' large, dynamic structures, new materials or materials precursors (such as membranes and molecular sieves) might be obtained. □

Received 16 December 1999; accepted 10 April 2000.

- Mayer, A. M. Floating magnets. *Nature* **18**, 258–260 (1878).
- Derr, L. A photographic study of Mayer's floating magnets. *Proc. Am. Acad. Arts Sci.* **44**, 525–528 (1909).
- Monckman, J. On the arrangement of electrified cylinders when attracted by an electrified sphere. *Proc. Camb. Phil. Soc.* **6**, 179–181 (1888).
- Schecter, D. A., Dubin, D. H. E., Fine, K. S. & Driscoll, C. F. Vortex crystals from 2D Euler flow: experiment and simulation. *Phys. Fluids* **11**, 905–914 (1999).
- Bubeck, R., Bechinger, C., Naser, S. & Leiderer, P. Melting and reentrant freezing of two-dimensional colloidal crystals in confined geometry. *Phys. Rev. Lett.* **82**, 3364–3367 (1999).
- Ghyka, M. *The Geometry of Art and Life* (Dover, New York, 1977).
- Ball, P. *The Self-Made Tapestry: Pattern Formation in Nature* (Oxford Univ. Press, New York, 1999).
- Whitesell, J. K. *Organised Molecular Assemblies in the Solid State* (Wiley, New York, 1999).
- Philp, D. & Stoddart, J. F. Self-assembly in natural and unnatural systems. *Angew. Chem. Int. Edn Engl.* **35**, 1155–1196 (1996).
- Lounasmaa, O. V. & Thuneberg, E. Vortices in rotating superfluid ³He. *Proc. Natl Acad. Sci. USA* **96**, 7760–7767 (1999).
- Bowden, N., Brittain, S., Evans, A. G., Hutchinson, J. W. & Whitesides, G. M. Spontaneous formation of ordered structures in thin films of metals supported on an elastomeric polymer. *Nature* **393**, 146–149 (1998).
- Bowden, N., Choi, I. S., Grzybowski, B. A. & Whitesides, G. M. Mesoscale self-assembly of hexagonal plates using lateral capillary forces: synthesis using the "capillary bond". *J. Am. Chem. Soc.* **121**, 5373–5391 (1999).
- Shinbrot, T. Competition between randomizing impacts and inelastic collisions in granular pattern formation. *Nature* **389**, 574–576 (1997).
- Burns, M. M., Fournier, J. M. & Golovchenko, J. A. Optical matter—crystallization and binding in intense optical fields. *Science* **249**, 749–754 (1990).
- Murray, C. A. & Grier, D. G. Colloidal crystals—solid particles suspended in fluid form ordered arrays with unusual and useful physical properties. *Am. Sci.* **83**, 238–245 (1995).
- Thompson, D'A. *On Growth and Form* (Dover, New York, 1992).
- Shapiro, J. A. Thinking about bacterial populations as multicellular organisms. *Annu. Rev. Microbiol.* **52**, 81–104 (1998).
- Berg, H. & Budrene, E. O. Dynamics of formation of symmetrical patterns by chemotactic bacteria. *Nature* **376**, 49–53 (1995).
- GomezLopez, M., Preece, J. A. & Stoddart, J. F. The art and science of self-assembling molecular

machines. *Nanotechnology* **7**, 183–192 (1996).

- Koschmieder, A. *Benard Cells and Taylor Vortices* (Cambridge Univ. Press, New York, 1993).
- Jakubith, S., Rotermund, H. H., Engel, W., von Oertzen, A. & Ertl, G. Spatiotemporal concentration patterns in a surface reaction: propagating and standing waves, rotating spirals, and turbulence. *Phys. Rev. Lett.* **65**, 3013–3016 (1990).
- Engelborghs, Y. Microtubules—dissipative structures formed by self-assembly. *Biosens. Bioelectron.* **9**, 685–689 (1994).
- Lindgren, K., Moore, C. & Nordahl, M. Complexity of two-dimensional patterns. *J. Stat. Phys.* **91**, 909–951 (1998).
- Singh, R., Maru, V. M. & Moharir, P. S. Complex chaotic systems and emergent phenomena. *J. Nonlinear Sci.* **8**, 235–259 (1998).
- Havelock, T. H. The stability of motion of rectilinear vortices in ring formation. *Phil. Mag.* **11**, 617–633 (1931).
- Morton, W. B. Vortex polygons. *Proc. R. Irish Acad. A* **42**, 21–29 (1935).
- Aref, H. & Vainchtein, D. L. Point vortices exhibit asymmetric equilibria. *Nature* **392**, 769–770 (1998).
- Thomson, J. J. *A Treatise on the Motion of Vortex Rings* (Macmillan, London, 1883).
- Dritschel, D. G. The stability and energetics of corotating uniform vortices. *J. Fluid Mech.* **157**, 95–134 (1985).
- Schonberg, J. A. & Hinch, E. J. Inertial migration of a sphere in Poiseuille flow. *J. Fluid Mech.* **203**, 517–524 (1989).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by DARPA and NSF.

Correspondence and requests for materials should be addressed to G.M.W. (e-mail: gwhitesides@gmwgroup.harvard.edu).

Structural basis for the fracture toughness of the shell of the conch *Strombus gigas*

S. Kamat*, X. Su*, R. Ballarini† & A. H. Heuer*

* Department of Materials Science and Engineering, † Department of Civil Engineering, Case Western Reserve University, Cleveland, Ohio 44106-7204, USA

Natural composite materials are renowned for their mechanical strength and toughness: despite being highly mineralized, with the organic component constituting not more than a few per cent of the composite material, the fracture toughness exceeds that of single crystals of the pure mineral by two to three orders of magnitude¹. The judicious placement of the organic matrix, relative to the mineral phase, and the hierarchical structural architecture extending over several distinct length scales both play crucial roles in the mechanical response of natural composites to external loads^{2–4}. Here we use transmission electron microscopy studies and beam bending experiments to show that the resistance of the shell of the conch *Strombus gigas* to catastrophic fracture can be understood quantitatively by invoking two energy-dissipating mechanisms: multiple microcracking in the outer layers at low mechanical loads, and crack bridging in the shell's tougher middle layers at higher loads. Both mechanisms are intimately associated with the so-called crossed lamellar microarchitecture of the shell, which provides for 'channel' cracking in the outer layers and uncracked structural features that bridge crack surfaces, thereby significantly increasing the work of fracture, and hence the toughness, of the material. Despite a high mineral content of about 99% (by volume) of aragonite, the shell of *Strombus gigas* can thus be considered a 'ceramic plywood', and can guide the biomimetic design of tough, lightweight structures.

The crossed lamellar shell of *Strombus gigas*, the giant pink Queen conch native to Caribbean habitats, contains structure at five

distinct length scales. Other mollusc shell microarchitectures are nacreous, foliated, prismatic, homogeneous, and complex crossed lamellar^{3,6}, but the crossed lamellar structure is the most common one, and is associated with the highest nominal fracture toughness, K_{Ic} (refs 7–10). This fracture-mechanics parameter measures the resistance of a solid to crack propagation, and is directly related to the state of stress near the tip of a crack at the onset of crack extension. For perfectly brittle materials, the surface energy created during fracture, 2γ , determines the energetics of the fracture process. The fracture toughness and the effective Young's modulus, E , are then related to γ through the Irwin relation $K_{Ic} = \sqrt{2\gamma E}$. Tougher materials exhibit additional energy-dissipating fracture mechanisms; for shells with the crossed lamellar structure, the additional energy required for crack propagation, J , arises because the crack surfaces are bridged by microstructural features resulting from the unique microarchitecture.

The hierarchy of the structural features that comprise the shell of *Strombus gigas*, and their characteristic dimensions, are shown in Fig. 1a, a highly schematic drawing summarizing our current knowledge of the structure. The basic building blocks are third-order, aragonite (a polymorph of calcium carbonate, CaCO_3) lamellae, which are about 100 nm by 250 nm in cross-section, many micrometres long, internally twinned, and completely surrounded by organic matrix. The coarsest features are macroscopic layers, which are best seen using scanning electron microscopy (SEM) of fractured surfaces of bend test specimens, Fig. 1b. Such specimens were taken from the last whorl of the shell, and contained three macroscopic layers. (The shell lip, which is present only in mature individuals, can contain more than three individual layers.) The higher-magnification images of the fracture surface, Fig. 1c and d, clearly show the first-order, second-order and third-order lamellae and their relative orientation, and fully justify the 'crossed lamellar' description of the microarchitecture.

Imaging of the organic matrix present in the lamellar interfaces in this heavily mineralized (~99% aragonitic CaCO_3) shell is

necessary to understand its mechanics. Figure 2a shows a partially demineralized polished section, in which first-order (BCCB, BCB) and second-order (CC) lamellar interfaces can clearly be seen. Figure 2b, c and d show transmission electron microscopy (TEM) images of first-order, second-order, and third-order lamellar interfaces, respectively. The judicious placement of the proteinaceous matrix surrounding and separating the several lamellar orders of the stiffer mineral phase profoundly affects the mechanical response of the shell.

During bending deformation, multiple channel cracks along first-order interfaces develop in the inner layer at low loads, but are arrested by the tough middle layer; catastrophic failure is mitigated until much higher loads. The mutual shielding produced by the interaction of these closely spaced cracks, which propagate along first-order lamellar interfaces, contributes to a higher stress for catastrophic failure (and in turn a higher work of fracture, which is given by the area under the stress-strain curve divided by the fracture surface area) compared to singly cracked specimens¹⁰. This aspect of the fracture response can be modelled¹¹ using a two-layer, elastically homogeneous structure with the fracture toughness of the middle layer, K_{Ic}^m , being λ times higher than that of the inner layer, K_{Ic}^i . For multiple cracks to develop under uniform tension, λ must be greater than 2. For λ as small as 3, a 20-fold increase in the work of fracture is achieved, together with an initiation stress that is three times higher than the stress required to fracture a specimen with a uniform toughness equal to that of the middle layer.

As the load is increased, a saturation channel crack density is eventually reached, and one or more of the channel cracks start to grow through the middle layer, along the interfaces between second-order lamellae. These cracks do not propagate catastrophically through the middle layer; instead they are retarded by the bridging action of the first-order lamellae. (Sometimes, they grow for a short distance along the interface between the inner and middle layers before traversing the middle layer.) As discussed next, this

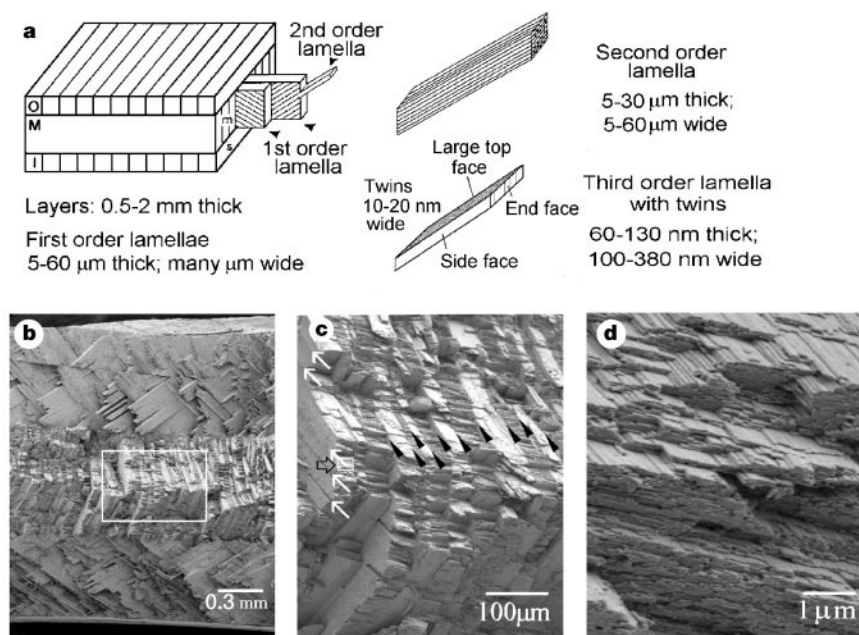


Figure 1 Microarchitecture of the shell of *Strombus gigas*. **a**, A schematic drawing of the crossed lamellar microarchitecture, including characteristic dimensions of the three lamellar orders. (O, M and I refer to outer, middle and inner (adjacent to the animal) layers, respectively.) **b**, **c** and **d**, SEM images of the fracture surface of a bend specimen taken at increasing magnification (the boxed areas in **b** and **c** (see open arrow in **c**) are shown in **c**

and **d**, respectively). The first- and second-order interfaces can be clearly seen and are indicated by black arrowheads and white arrows, respectively, in **c**. Third-order lamellae are readily observed in **d**. The organic matrix cannot be discerned in such images; visualizing the matrix requires transmission electron microscopy (see Fig. 2b–d).

phenomenon is responsible for ultimate load and work of fracture values that are significantly higher than those that would result from the propagation of unbridged cracks (as assumed by the simple two-layer linear model).

Damage associated with catastrophic fracture is shown in Fig. 3. A surface inclined to a strength-defining crack within a fractured specimen is shown in Fig. 3a; the geometry is sketched in Fig. 3b. As the crack front cleaves the interfaces between second-order lamellae, every other first-order lamella remains intact, and contributes to the reduction of the crack opening displacement (COD). The net bridging tractions on the strength-defining crack are influenced by the stiffness of the first-order lamellae, as well as by the frictional stress along the interfacial cracks that develop along the parallel interfaces between first-order lamellae.

Figure 3c and d shows microcracks within the middle layer of a fractured, dye-impregnated sample; the micrograph was obtained by polishing away the outer layer. The load–displacement curve for this specimen, shown in Fig. 3e, clearly illustrates the onset of channel cracking, interlayer delamination, and the high work of fracture of the shell.

Notched fracture-mechanics bend specimens were also prepared, as shown in Fig. 4a. For initial machined notches confined to the inner layer, the crack front propagated self-similarly along the first-order lamellar interfaces. For those samples containing notches machined into the middle layer, cracks propagated directly along the second order interfaces at $\sim 45^\circ$ to the applied flexural tensile

stress. (Some specimens developed short delamination cracks at the interface between the inner and middle layers.) The inset in Fig. 4a presents the nominal fracture toughness and its standard deviation derived from these tests. For initial notches in the inner layer, these values represent the toughness of the protein interface; for those in the middle layer, they represent an effective initiation toughness of the middle layer.

We note two points about these data: the negligible difference in interface toughness between wet and dry specimens, and the approximately four times difference in toughness between the inner and middle layers. This last point is consistent with the requirement derived from the multiple cracking model¹¹; $\lambda > 2$ results in multiple channel cracks. However, the multiple channel-cracking model cannot predict the ultimate load and work of fracture (and their dependence on shell dimensions), because crack propagation in the middle layer is not accurately characterized by a constant K_{Ic}^m . Instead, it is associated with a form of crack bridging that is analogous to fibre bridging in fibrous composites, as we now show.

Load–displacement curves for specimens in which the cracks extended into the middle layers can be used to calibrate a micro-mechanics crack-bridging model, which can then be used to predict the effects of shell dimensions on ultimate load and work of fracture. (Previous work^{8–10} contained critical nominal stresses and work of fracture for varying shell dimensions, with no rationally derived relationships that could generate structural predictions.) Such

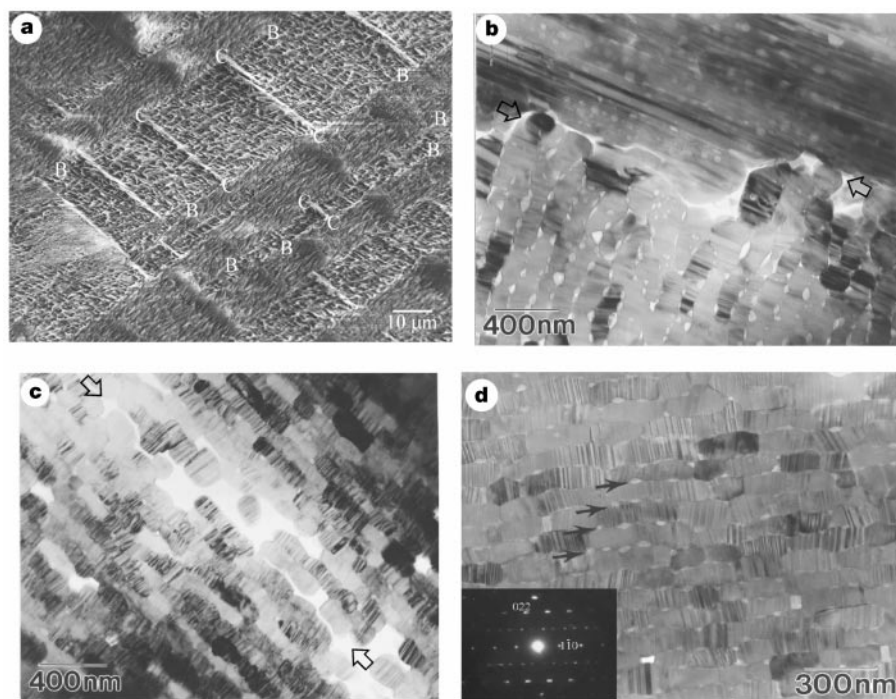


Figure 2 Images of shell specimens that show lamellar interfaces. **a**, SEM image of a polished, partially demineralized shell section showing interfaces separating second-order lamellae (CC) and interfaces separating first-order lamellae (BCCB and BCB). The long axes of the third-order lamellae are at an angle of about 45° to this polished section. These images confirm previous claims⁸ that the organic matrix surrounds every crystallite and is present at every interface. However, the boundaries of the second-order lamellae are thicker than the boundaries separating first-order lamellae. Within any second-order lamella, protein matrix associated with third-order lamellae can also be seen. **b**, **c**, TEM micrographs of ion-thinned specimens showing a first-order interface in **b**, and a second-order interface in **c** (the arrowed feature in each image). The interface thickness varies from 10 to 220 nm for the first-order interfaces and 20 to 320 nm for the second-order interfaces. It is clear that adjacent first-order lamellae are misoriented by $\sim 90^\circ$, and that

the side faces of the third-order lamellae (see Fig. 1a) are parallel to the first-order interfaces. On the other hand, adjacent second-order lamellae are not misoriented; rather, the interfaces between adjacent lamellae are thicker regions of matrix, even thicker than the first-order planar interfaces. At the scale visible in the TEM, first-order lamellar interfaces are neither flat nor planar. **d**, TEM end-face image, also of an ion-thinned specimen. The protein sheaths present between third-order lamellae are made up of thin membrane-like organic sheets connected to thicker globular proteins. (That these were proteinaceous and not empty space was confirmed by chemical microanalysis using electron energy-loss spectroscopy, EELS). The globular features are arranged along the top and bottom of each third-order lamella but not along the sides of these lamellae. The inset diffraction pattern reveals that the vertical features present in every third-order lamella are 'polysynthetic' growth twins.

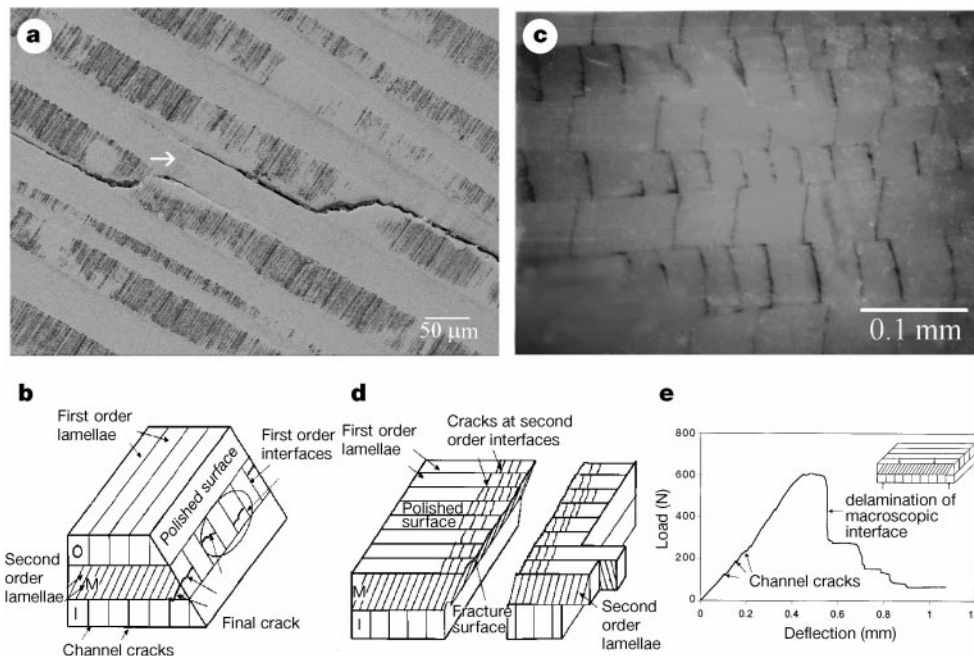


Figure 3 Damage development in the shell of *Strombus gigas* due to the difficulties of crack propagation within the middle layer. **a**, SEM image of a fractured specimen that underwent graceful, that is, non-catastrophic, failure and was intact after removal from the testing machine; a plane inclined to that of the strength-defining crack was then polished. **b**, Schematic drawing that shows that the crack penetrated every other first-order lamella, propagating long distances between the favourably and unfavourably oriented lamellae. In the arrowed region in **a**, the crack was unable to penetrate the first-order lamella and

caused delamination at the first-order interface. **c**, Optical micrograph (reflected light) showing stained microcracks within the middle layer of a failed sample near the final crack after dye penetration and serial sectioning, while **d** shows the fracture geometry. This microcracking within the middle layer accompanying final failure was restricted to a region ~1 mm from the plane of the final crack. **e**, Load-deflection curve for the specimen of **c**. The occurrence of channel cracks and layer delamination is evident in the load-deflection curve.

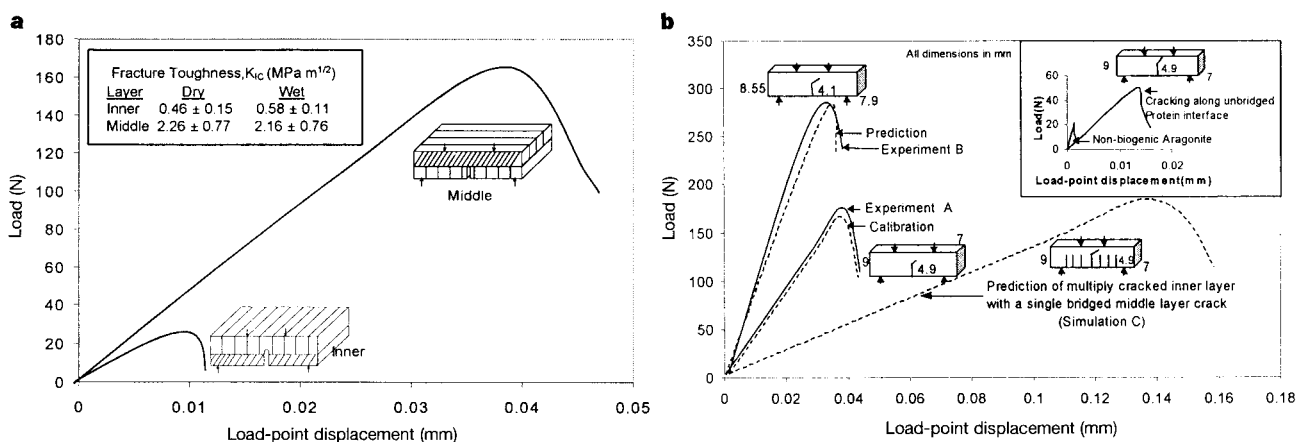


Figure 4 Load-displacement curves of notched bend test specimens. **a**, Curves measured on test specimens where the initial notch tip was machined either into the inner layer or into the middle layer. The inset in **a** shows the average fracture toughness K_{Ic} and its standard deviation, calculated from the peak loads for dry samples and for those tested wet after storing in artificial sea water for 30 days. **b**, Experimental and theoretical load-deflection curves for specimens containing middle layer notches. The 'calibration' sample A allows determination of the parameters of the crack bridging model, and suggests that $\beta = 630 \text{ N mm}^{-5/2}$ and $\text{COD}_c = 5 \text{ } \mu\text{m}$. These parameters allow good prediction of the load-deflection response for a specimen (B) with different geometry. Comparing the theoretical

unnotched beam response (simulation C, involving multiple cracking followed by a single channel crack propagating through the bridged middle layer) with the theoretically predicted response of beams comprised solely of non-biogenic aragonite ($K_{Ic} = 0.25 \text{ MPa m}^{1/2}$) or a beam in which cracks propagate unbridged along a mineral-protein interface ($K_{Ic} = 0.6 \text{ MPa m}^{1/2}$) shown in the inset clearly illustrates the impressive fracture toughness of the shell of *Strombus gigas*. In the crack-bridging model, equilibrium configurations are calculated under the assumption that the strength-defining crack is bridged according to the previously defined cohesive tractions, and extends at a stress intensity factor equal to the toughness of the mineral-protein interface.

nonlinear fracture mechanics theories¹² approximate, through tractions that resist crack opening, the net effects of various bridging mechanisms, such as bridging and frictional sliding between first-order lamellae and microcracking. Our micromechanical model predicts that the magnitudes of the bridging stresses, p , are proportional to the square root of the COD ($p = \beta\sqrt{\text{COD}}$), and are operative up to a critical opening COD_{cr} ; here, β incorporates the effects of all possible energy-dissipating mechanisms. The additional energy associated with a fully developed bridging zone (whose tail has reached COD_{cr}) is given as:

$$J = \int_0^{\text{COD}_{\text{cr}}} p(\text{COD})d(\text{COD}) = \frac{2}{3}\beta(\text{COD}_{\text{cr}})^{3/2}$$

We note that in this type of theory, which has been successful with other brittle matrix composites¹², the bridging law is the true material property.

Figure 4b shows load–displacement curves for a notched specimen (experiment A) and the ‘calibration’ simulation; a good fit to experiment is obtained using $\beta = 630 \text{ N m}^{-5/2}$ and $\text{COD}_{\text{cr}} = 5 \mu\text{m}$ ($J = 0.15 \text{ N mm}^{-1}$). A comparison between the predictions of the now-calibrated bridging model and the experimental response of a notched beam of different dimensions (experiment B) is also shown in Fig. 4b. The model accurately predicts the behaviour of cracks propagating through the middle layer, and in turn, the effects of shell dimensions. The model can also predict the response of a multiply cracked unnotched shell sample, where final fracture results from the (stochastic) propagation through the middle layer of a single channel crack (simulation C). The load–deflection curve for this case suggests that the combination of multiple cracking and ‘fibre’ bridging increases the work of fracture by an additional order of magnitude.

A comparison of this last simulation with the mechanical response predicted for a crack growing self-similarly in a pure aragonite shell, or for a crack growing at 45° along an unbridged protein interface (see the inset of Fig. 4b) emphasizes the truly impressive fracture toughness of the shell of *Strombus gigas*. We consider that the microarchitecture of the *Strombus gigas* shell—that is, ‘ceramic plywood’—could guide the design of lightweight structural composites. □

Received 28 October 1999; accepted 8 May 1999.

- Wegst, U. G. K. & Ashby, M. F. Material selection charts for natural materials. Report No. CUED/C_EDC/TR55 (Cambridge Univ., 1997).
- Taylor, J. D. & Layman, M. The mechanical properties of bivalve (mollusca) shell structures. *Palaeontology* **15**, 73–86 (1972).
- Currey, J. D. & Taylor, J. D. The mechanical behaviour of some molluscan hard tissue. *J. Zool. Lond.* **173**, 395–406 (1974).
- Currey, J. D. in *Mechanical Properties of Biological Materials* (eds Currey, J. D. & Vincent, J. F. V.) 75–97 (Cambridge Univ. Press, 1980).
- Bogild, O. B. The shell structure of the mollusks. *K. Danske Vidensk. Selsk. Skr.* **2**, 232–325 (1930).
- Carter, J. G. in *Skeletal Growth of Aquatic Organisms* (eds Rhoads, D. C. & Lutz, R. A.) 69–113 (Plenum, New York, 1980).
- Curry, J. D. & Kohn, A. J. Fracture in the crossed-lamellar structure of *Conus* shells. *J. Mater. Sci.* **11**, 1615–1623 (1976).
- Jackson, A. P., Vincent, J. F. V. & Turner, R. M. The mechanical design of nacre. *Proc. R. Soc. Lond. B* **234**, 415–440 (1988).
- Jackson, A. P., Vincent, J. F. V. & Turner, R. M. Comparison of nacre with other ceramic composites. *J. Mater. Sci.* **25**, 3173–3178 (1990).
- Kuhn-Spearing, L. T., Kessler, H., Spearing, S. M., Ballarini, R. & Heuer, A. H. Fracture mechanisms of the *Strombus gigas* conch shell: Implication for the design of brittle laminates. *J. Mater. Sci.* **31**, 6583–6594 (1996).
- Kessler, H., Ballarini, R., Mullen, R. L., Kuhn, L. T. & Heuer, A. H. A biomimetic example of brittle toughening: (I) steady state multiple cracking. *Comput. Mater. Sci.* **5**, 157–166 (1996).
- Cox, B. & Marshall, D. B. Concepts for bridged cracks in fracture and fatigue. Overview No. 111. *Acta Metall. Mater.* **42**, 341–363 (1994).

Acknowledgements

A.H.H. and R.B. thank V. Laria, H. Kessler and L. Kuhn, former postdoctoral research fellows, for their contributions to his understanding of the shell of *Strombus gigas*. This research was supported by EPRI.

Correspondence and requests for materials should be addressed to A. H. H. (e-mail: ahh@po.cwru.edu).

Palaeotemperature reconstruction from noble gases in ground water taking into account equilibration with entrapped air

W. Aeschbach-Hertig[†], F. Peeters^{*}, U. Beyerle^{*} & R. Kipfer[†]

^{*} Department of Water Resources and Drinking Water, Swiss Federal Institute of Environmental Science and Technology (EAWAG), CH-8600 Dübendorf, Switzerland

[†] Environmental Physics, [‡] Isotope Geology, Swiss Federal Institute of Technology (ETH), CH-8902 Zurich, Switzerland

Noble-gas concentrations in ground water have been used as a proxy for past air temperatures^{1–7}, but the accuracy of this approach has been limited by the existence of a temperature-independent component of the noble gases in ground water, termed ‘excess air’, whose origin and composition is poorly understood^{7–9}. In particular, the evidence from noble gases in a Brazilian aquifer for a cooling of more than 5°C in tropical America during the Last Glacial Maximum⁴ has been called into question⁹. Here we propose a model for dissolved gases in ground water, which describes the formation of excess air by equilibration of ground water with entrapped air in quasi-saturated soils^{10–12}. Our model predicts previously unexplained noble-gas data sets, including the concentration of atmospheric helium, and yields consistent results for the non-atmospheric helium isotopes that are used for dating ground water. Using this model of excess air, we re-evaluate the use of noble gases from ground water for reconstructing past temperatures. Our results corroborate the inferred cooling in Brazil during the Last Glacial Maximum⁴, and indicate that even larger cooling took place at mid-latitudes.

Noble gases in ground water consist of three components: (1) dissolved air at solubility equilibrium, (2) certain isotopes from radioactive decay¹³, and (3) ‘excess air’⁸. The temperature dependence of the first component has been used to infer recharge temperatures of ground water in order to reconstruct palaeotemperatures^{1–7}. The second component is of importance for He (³He from ³H, ⁴He from U/Th), and has been used extensively for groundwater dating^{14–19}. Little is known about the origin and composition of the third component, although excess air may contain information about the environmental conditions pertaining during infiltration^{7,20–22}. An understanding of the excess-air phenomenon is needed for the reliable calculation of noble-gas temperatures (NGTs) and groundwater ages.

Usually it is assumed that excess air is formed by total dissolution

Table 1 Statistical analysis of fits of excess-air models to noble-gas data sets

Data set	N	TD model		PR model		CE model	
		χ^2	P	χ^2	P	χ^2	P
Brazil	20	626.4	$< 10^{-14}$	71.8	9×10^{-8}	35.8	0.016
Oman	9	153.0	$< 10^{-14}$	55.2	1×10^{-8}	20.4	0.016
Maryland	20	246.8	$< 10^{-14}$	44.5	0.001	18.2	0.574
Belgium	28	303.6	$< 10^{-14}$	55.7	0.001	12.8	0.994
Belgium He*	28	358.6	$< 10^{-14}$	191.5	1×10^{-14}	34.2	0.990

N is the number of samples of each data set. χ^2 is the sum of the weighted squared deviations between modelled and measured noble-gas concentrations, summed over all samples of a data set²³. The expected value of χ^2 is equal to the number of degrees of freedom, which is the number of fitted concentrations ($4N$, with He $5N$) minus the number of free model parameters ($2N$ (T , A_{at}) for the TD model, $3N$ (T , A_{at} , R or T , A_{at} , F) otherwise). P is the probability that χ^2 exceeds the observed value. Models with $P < 0.01$ are rejected.

* Belgian data set with the calculated He_{atm} (Methods, Fig. 1) as additional constraint.

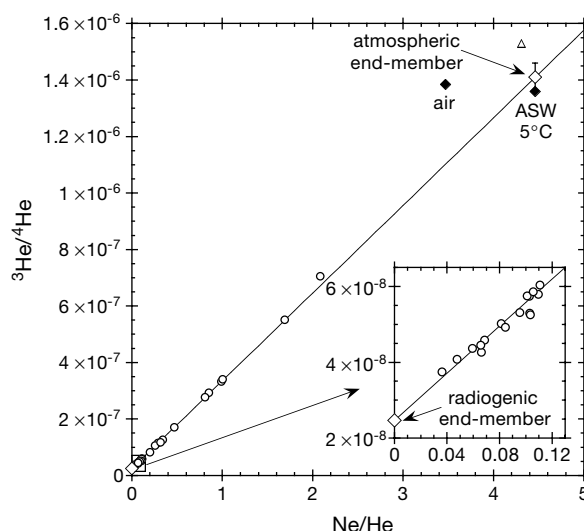


Figure 1 He isotope ratio versus Ne/He elemental ratio of samples from Belgium. The data define a mixing line between a radiogenic and an atmospheric end-member. The linear regression was calculated by a least-squares fit weighted with both x and y errors (errors are similar to the size of the symbols). One sample (triangle) from a shallow well in the

of small air bubbles trapped in soil pores^{2,3,5,8}. In this total dissolution (TD) model, the concentrations of dissolved atmospheric gases are²³:

$$C_i(T, S, P, A_d) = C_i^*(T, S, P) + A_d z_i \quad (1)$$

where $C_i^*(T, S, P)$ are the moist-air solubility equilibrium concentrations as functions of temperature, salinity, and atmospheric pressure, A_d is the concentration of totally dissolved dry air, and z_i are the volume fractions of the individual gases in dry air. However, because the TD model does not yield consistent NGTs for the noble-gas record from Brazil, a model of excess-air fractionation by partial diffusive re-equilibration (PR model) was introduced⁴, which may be written as²³:

$$C_i(T, S, P, A_d, R) = C_i^*(T, S, P) + A_d z_i e^{-R \frac{D_i}{D_{Ne}}} \quad (2)$$

where R is a parameter describing the degree of re-equilibration and D_i are the molecular diffusion coefficients. Yet, even the PR model does not provide an adequate description of the Brazilian noble-gas data⁹.

Here we propose a new model explaining excess air as the result of equilibration of ground water with persistent entrapped air. The occurrence of air entrapment during groundwater infiltration is well known^{10–12}, but has never been quantitatively linked to the phenomenon of excess air. The basic assumption of our model is that solubility equilibrium is attained in a closed system of initially air-saturated water and a finite volume of entrapped air under a

recharge area appears to be affected by ^3He from the decay of bomb tritium, and was thus excluded from the fit. The point for air-saturated water (ASW) is shown for the representative temperature of 5 °C.

constant hydrostatic pressure. The equations of this closed-system equilibration (CE) model are:

$$C_i(T, S, P, A_e, F) = C_i^*(T, S, P) + \frac{(1 - F)A_e z_i}{1 + FA_e z_i / C_i^*} \quad (3)$$

where A_e is the initial amount of entrapped air per unit mass of water and F is the fractionation parameter (Methods). Note that A_e (entrapped air) is the same as A_d (dissolved air) only in the case of total dissolution. F is the ratio of two parameters with a clear physical meaning: v , the ratio of the entrapped gas volumes in the final and initial state, and q , the ratio of the dry gas pressure in the trapped gas to that in the free atmosphere. The values of v and q can be calculated from $F = v/q$ and A_e , using the condition that the sum of all partial pressures in the trapped volume equals the total pressure. Equation (3) describes all possible cases between no excess air ($F = v = q = 1$), pure excess air ($F = 0$, $v = 0$ or q infinite), and a pure pressure effect, that is, equilibration with the atmosphere at increased pressure ($v = 1$, $q > 1$, $F = 1/q < 1$, A_e infinite). Thus, it is a very general description of the concentrations of dissolved gases in ground water.

Equations (1) to (3) describe the concentrations of Ne, Ar, Kr and Xe (He is usually affected by non-atmospheric sources) with up to five parameters. In practice, S and P are usually known ($S \approx 0$ for meteoric water, and P can be calculated from the altitude of the recharge area). The remaining unknown parameters have traditionally been determined by iteration^{3,4,7}. Recently, weighted least-squares techniques to invert the model equations have been developed, enabling quantitative assessment of the ability of different conceptual models to explain the observations^{9,23}.

We checked the practical applicability of the models by fitting them to two noble-gas data sets from tropical regions, northeastern Brazil⁴ and Oman⁶, and two from temperate regions, Maryland (USA)²⁴ and Belgium. The goodness of fit was quantified by applying the χ^2 test to the ensemble of the samples of each data set^{9,23} (Table 1). Although the PR model performs far better than the TD model, it does not describe the observed noble-gas concentrations within their experimental uncertainty. In all cases, only the CE model yields an acceptable description of the data. The physical plausibility of the CE model can be checked by comparing the results obtained for its physically interpretable parameters with empirical findings on entrapped air and excess air (Table 2). The

Table 2 Values of excess-air parameters obtained from fits of the CE model

Data set	A_e	v	q	F
Brazil	0.017 ± 0.005	0.53 ± 0.19	1.56 ± 0.29	0.37 ± 0.17
Oman	0.025 ± 0.020	0.78 ± 0.12	1.31 ± 0.19	0.61 ± 0.14
Maryland	0.049 ± 0.059	0.88 ± 0.07	1.16 ± 0.04	0.76 ± 0.06
Belgium	0.083 ± 0.074	0.91 ± 0.08	1.16 ± 0.08	0.79 ± 0.09
Expected*	0.02–0.18	~0.9	~1.2	~0.75

A_e is given in $\text{cm}^3 \text{STP g}^{-1}$; the other parameters are dimensionless. Mean values and standard deviations of the results from the individual samples are given.

* Expected values were derived from empirical field data as follows. A_e : entrapped air was found to occupy between 2% and 15% of the pore space (corresponding to 2–18% of water volume) during groundwater infiltration¹¹. v : Typical concentrations of dissolved excess air are of the order of 0.001 to 0.01 $\text{cm}^3 \text{STP g}^{-1}$ (10–100% ΔNe)^{5,7,23}, indicating dissolution of only about 10% of A_e , and hence corresponding to v -values of about 0.9. q : air entrapment occurs in the uppermost metres of the aquifer, where the water table fluctuates^{11,12}. q increases above 1 by about 0.1 per metre of hydrostatic overload. F : from $F = v/q$.

Table 3 Model dependence of non-atmospheric He components and corresponding ages

Model	Maryland: 4 young samples		Switzerland: 5 fractionated samples		
	$^4\text{He}_{\text{rad}}$ ($10^{-9} \text{ cm}^3 \text{ STP g}^{-1}$)	^4He age (yr)	$^4\text{He}_{\text{rad}}$ ($10^{-9} \text{ cm}^3 \text{ STP g}^{-1}$)	$^3\text{He}_{\text{tri}}$ ($10^{-15} \text{ cm}^3 \text{ STP g}^{-1}$)	$^3\text{H}-^3\text{He}$ age (yr)
TD model	-1.0 ± 0.8	-630 ± 490	-0.5 ± 0.8	-0.1 ± 1.4	0.0 ± 0.4
PR model	7.8 ± 2.3	$4,700 \pm 1,400$	5.9 ± 1.2	9.4 ± 2.0	2.4 ± 0.5
CE model	0.3 ± 0.5	210 ± 290	0.6 ± 0.9	1.4 ± 1.4	0.4 ± 0.4
Expected*	< 0.08	< 50	0.8 ± 0.4	3.1 ± 1.3	0.8 ± 0.3

Mean values and standard deviations of the results from the individual samples are given.

* Expected values were derived as follows. Maryland: the 4 samples discussed here are from shallow wells in the recharge area and must be less than 50 yr old according to ^3H and ^3He data. Conversion between $^4\text{He}_{\text{rad}}$ and ages is done using the *in situ* production rate of $1.6 \times 10^{-12} \text{ cm}^3 \text{ STP g}^{-1} \text{ yr}^{-1}$, derived from U and Th analyses on sediment samples and supported by ^{14}C ages²⁴. Switzerland: the 5 samples discussed here, all from one particular sampling date, are the only ones with fractionated excess air out of a data set of 48 samples^{17,25}. Expected results were estimated based on 10 samples from the same boreholes but other sampling dates.

data sets from Maryland and Belgium yield values within the expected range. The two tropical records yield higher values of q and lower values of A_e and v . This finding may reflect actual differences in the infiltration regime, for example, larger water-table fluctuations (higher q) in the tropical aquifers.

The conclusion that only the CE model provides a realistic description of the data is further supported by an analysis of the usually neglected He data. Because of its low solubility and high diffusivity, He reacts most sensitively to excess air and to diffusive fractionation. However, most palaeogroundwaters contain large excesses of non-atmospheric, usually radiogenic, He. Atmospheric (He_{atm}) and radiogenic (He_{rad}) He components may be separated based on their very different $^3\text{He}/^4\text{He}$ ratios²⁵. Our data from Belgium define a two-component mixing line between a radiogenic and an atmospheric end-member (Fig. 1). Based on the position along this line, each sample was split into its components (Methods).

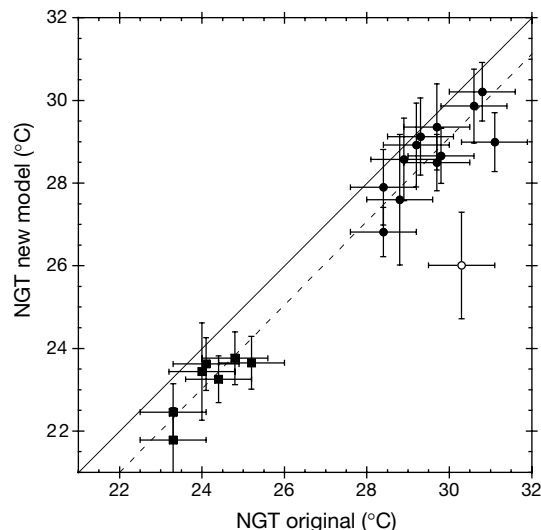


Figure 2 Comparison of new and original noble gas temperatures (NGTs) from Brazil. In the original publication⁴, NGTs were calculated using an iterative technique to correct for fractionated excess air according to the PR model. We re-evaluated NGTs by inversion of the new CE model for fractionated excess air. Symbols reflect the clusters of samples identified in the original work: filled circles correspond to high temperatures and low ^{14}C -ages, filled squares to lower temperatures and higher ages, and the open circle refers to a special sample (no. 17) which for several reasons was omitted in calculating the glacial–interglacial temperature difference⁴. The regression line (dashed) through the data (excluding sample 17) is practically parallel to the 1:1 line (solid), and indicates that the new NGTs are on average 1 °C lower than the original values. The mean new NGT of the warm cluster is 28.7 ± 0.3 °C (compared to 29.6 °C in the original evaluation) in good agreement with the mean local ground temperature of 29.1 °C. The cool cluster yields a mean NGT of 23.1 ± 0.3 °C (original value 24.2 °C). We note that temperature difference between warm and cold clusters remains practically unchanged (5.6 ± 0.4 °C).

The calculated He_{atm} was then used as a fifth, excess air sensitive, constraint in the inversion of equations (1) to (3). With this additional constraint, the CE model still provides a good fit, whereas the PR model becomes clearly unacceptable (Table 1).

The predictions of the PR and CE models for He_{atm} differ strongly due to high diffusive He loss in the PR model. This fact has consequences for groundwater dating based on the accumulation of radiogenic $^4\text{He}_{\text{rad}}$ and tritiogenic $^3\text{He}_{\text{tri}}$. These He components are best determined by measuring all noble gases, and using the parameters derived from the inverse modelling of the heavier noble gases to calculate He_{atm} . We used this approach to calculate $^4\text{He}_{\text{rad}}$ in samples from the recharge area of the coastal aquifer in Maryland²⁴ as well as $^4\text{He}_{\text{rad}}$ and $^3\text{He}_{\text{tri}}$ in samples from a shallow alluvial aquifer in Switzerland^{17,23}. In both examples, we can estimate the expected results (Table 3). The TD model implies obviously inconsistent negative values of the non-atmospheric He components and corresponding ages. The PR model yields unrealistically high ages. Only the results obtained with the CE model are in accordance with expectations.

For dating with He isotopes, usually only Ne is measured to calculate He_{atm} , based on the assumption that excess air has the atmospheric He/Ne ratio of 0.288 (TD model)^{14–17}. As in our examples, this assumption sometimes leads to negative ages^{16,17}. This inconsistency can be resolved by assuming fractionated excess air with a lower He/Ne ratio. In contrast to the PR model, the CE model predicts a lower limit of the excess air He/Ne ratio, given by the value for equilibrated water (0.22 to 0.25, depending on temperature). This value can be used to define an upper boundary for non-atmospheric He components and related ages. Our results highlight the advantage of combining analyses of He isotopes and atmospheric noble gases. He data may be useful to identify excess air fractionation and hence to obtain more reliable NGTs. Conversely, atmospheric noble-gas data provide a firm basis for the calculation of the non-atmospheric ^4He and ^3He components needed for dating.

The most important implications of our excess air model concern the reliability of noble-gas palaeoclimate records, particularly with regard to the controversial issue of tropical temperatures during the Last Glacial Maximum. The discrepancy between weak glacial cooling (≤ 2 °C) indicated by most oceanic palaeoclimate proxies^{26,27}, and a strong temperature change (~ 5 °C) indicated by continental records^{4,28,29}, is still not entirely explained.

The reliability of the important evidence for a large continental cooling provided by the Brazilian noble-gas study depends on an appropriate understanding of excess air. The original interpretation was based on the PR model⁴, which is inconsistent with the data⁹ (Table 1). Applying the CE model and the inverse fitting procedure²³, we obtain consistent new NGTs that are systematically lower than the original values by about 1 °C (Fig. 2). The data form two clusters, one corresponding to high temperatures and low ^{14}C ages, the other to lower temperatures and higher ages. The temperature difference between the clusters is 5.6 ± 0.4 °C, indistin-

guishable from the original result of $5.4 \pm 0.6^\circ\text{C}$. Our internally consistent and mathematically rigorous analysis of the data reinforces the conclusion of a $\sim 5^\circ\text{C}$ glacial–interglacial temperature difference in tropical Brazil. The hypothesis of a large tropical cooling is further supported by recent noble-gas records indicating a glacial cooling of 6.5°C in Oman⁶ and 5.3°C in Namibia²². Whereas the Oman data could only be interpreted with the new CE model, the results from Namibia were based on the PR model and may need to be re-evaluated.

The CE model is crucial for a consistent interpretation not only of the tropical noble-gas records, but also those from Maryland and Belgium (Table 1). NGTs calculated with the CE model suggest an even larger ($7\text{--}9^\circ\text{C}$) glacial cooling for these mid-latitude records. The original interpretation of other noble-gas data sets from temperate regions was based on the TD model and has also been questioned⁹. We expect that a re-evaluation of these data will primarily affect the individual NGTs rather than the reconstructed glacial–interglacial temperature difference, as exemplified by the Brazilian record. The inverse technique of calculating NGTs^{9,23} and the view of excess air described here provide a mathematically and physically sound theoretical foundation for the noble-gas palaeothermometer. The resulting reliable NGTs may be used to calibrate the ^{18}O palaeothermometer by simultaneous analysis of noble gases and ^{18}O in aquifers⁵.

Finally, our model provides a link between investigations of dissolved gases in ground water and the study of air entrapment in quasi-saturated soils. We are currently investigating relationships between entrapped air, excess air, fractionation, and infiltration conditions in laboratory and field studies. Eventually, excess air may provide information just as important as that derived from the other noble-gas components in ground water. □

Methods

Model derivation

According to Henry's law, at solubility equilibrium the gas concentrations C_i in solution are proportional to the partial pressures p_i in the gas phase:

$$p_i = H_i(T, S)C_i \quad (4)$$

where $H_i(T, S)$ is the Henry coefficient, depending on temperature T and salinity S .

We describe a closed system consisting of a water volume V_w and an initial trapped air volume V_g^0 under a constant total pressure $P_g = P + P_h$, where P is the atmospheric and P_h the hydrostatic pressure. In the initial state, the dissolved gas concentrations are in atmospheric solubility equilibrium²³:

$$C_i^*(T, S, P) = \frac{p_i^{\text{atm}}}{H_i(T, S)} = \frac{(P - e(T))z_i}{H_i(T, S)} \quad (5)$$

where $e(T)$ is the saturation water vapour pressure and z_i are the volume fractions in dry air. The volume at STP of dry entrapped air per unit mass of water is:

$$A_e = \frac{V_g^0}{\rho(T, S)V_w} \frac{(P_g - e(T))}{P_0} \quad (6)$$

where $\rho(T, S)$ is the water density, and P_0 is the standard pressure (1 atm).

In the final state, a gas volume V_g remains and equilibrium is achieved, that is, all gases partition between V_g and V_w according to Henry's law, equation (4). In a closed system, the total gas amounts must be conserved, that is, the sum of the number of moles of each gas in the water and the gas phase is equal in the initial and the final state.

$$n_{i,w}^0 + n_{i,g}^0 = n_{i,w} + n_{i,g} \quad (7)$$

Inserting equations (4) to (6) into (7) and using the ideal-gas law leads to equations that describe the concentrations of dissolved gases in the water in the final state:

$$C_i = (C_i^* + A_e z_i) \left(1 + \frac{V_g (P - e) z_i}{V_w \rho P_0 C_i^*} \right)^{-1} \quad (8)$$

where $C_i^* = C_i^*(T, S, P)$ are the moist air solubility equilibrium concentrations as in equation (5). By defining $v = V_g/V_g^0$, $q = (P_g - e)/(P - e)$, and $F = v/q$, we finally arrive at the model equations (3).

Separation of He components

Because the radiogenic He component (He_{rad}) is highly variable and can be overwhel-

mingly large, the measured concentrations of ^4He or total He (He_{meas}) provide no information on the atmospheric He component (He_{atm}) to be used to invert the model equations (1) to (3). However, because the typical radiogenic $^3\text{He}/^4\text{He}$ ratio ($R_{\text{rad}} \approx 2 \times 10^{-8}$) is two orders of magnitude smaller than that of the atmospheric sub-components (air: $R_a = 1.384 \times 10^{-6}$, air-saturated water: $R_{\text{asw}} \approx 0.983 R_a$; ref. 30), the radiogenic component is much less dominant for ^3He . If He_{meas} in a given aquifer results from a simple two-component mixture of He_{atm} and He_{rad} , both with uniform isotopic compositions R_{atm} and R_{rad} , measuring both He isotopes yields information on He_{atm} . To extract this information, we have to use the ensemble of the data set.

A plot of the measured $^3\text{He}/^4\text{He}$ isotope ratio versus the Ne/He elemental ratio can serve both as a check of the two-component mixing hypothesis as well as to define the isotopic composition of the end-members. The excellent linear correlation in our example from Belgium (Fig. 1) confirms that each sample is essentially a mixture of two end-members. The intercept of the regression line (calculated by a least-squares fit weighted with both x and y errors) at Ne/He = 0 defines $R_{\text{rad}} = (2.47 \pm 0.20) \times 10^{-8}$, typical for radiogenic He. The atmospheric end-member must lie on the regression line in the vicinity of the point for air-saturated water (ASW). Addition of ^3He from the decay of natural (pre-bomb) tritium shifts the end-member slightly upwards. We take the $^3\text{He}/^4\text{He}$ ratio defined by the regression line at the Ne/He value of the ASW point as the best estimate for R_{atm} , with an uncertainty large enough to include R_{asw} , yielding $R_{\text{atm}} = (1.41 \pm 0.05) \times 10^{-6}$.

The measured He isotope ratios (R_{meas}) reflect the mixing ratios of the two end-members. Using R_{atm} and R_{rad} defined from the entire data set, the contribution of He_{atm} to each individual sample can be calculated:

$$\text{He}_{\text{atm}} = \text{He}_{\text{meas}} \frac{(R_{\text{meas}} - R_{\text{rad}})}{(R_{\text{atm}} - R_{\text{rad}})} \quad (9)$$

In the Belgian data set, the uncertainty of the calculated He_{atm} ranges from 4% (due to the uncertainty in R_{atm}) up to 15% for the samples with the highest He_{rad} . Even this limited precision is however sufficient to distinguish between the PR and CE models, because their predictions for He_{atm} are very different (Table 1).

Received 14 October 1999; accepted 18 May 2000.

- Mazor, E. Paleotemperatures and other hydrological parameters deduced from gases dissolved in groundwaters, Jordan Rift Valley, Israel. *Geochim. Cosmochim. Acta* **36**, 1321–1336 (1972).
- Andrews, J. N. & Lee, D. J. Inert gases in groundwater from the Bunter Sandstone of England as indicators of age and palaeoclimatic trends. *J. Hydrol.* **41**, 233–252 (1979).
- Stute, M. & Schlosser, P. in *Climate Change in Continental Isotopic Records* (eds Swart, P. K., Lohmann, K. C., McKenzie, J. & Savin, S.) 89–100 (American Geophysical Union, Washington DC, 1993).
- Stute, M. *et al.* Cooling of tropical Brazil (5°C) during the Last Glacial Maximum. *Science* **269**, 379–383 (1995).
- Beyerle, U. *et al.* Climate and groundwater recharge during the last glaciation in an ice-covered region. *Science* **282**, 731–734 (1998).
- Weyhenmeyer, C. E. *et al.* Cool glacial temperatures and changes in moisture source recorded in Oman groundwaters. *Science* **287**, 842–845 (2000).
- Stute, M. & Schlosser, P. in *Environmental Tracers in Subsurface Hydrology* (eds Cook, P. & Herczeg, A. L.) 349–377 (Kluwer Academic, Boston, 2000).
- Heaton, T. H. E. & Vogel, J. C. "Excess air" in groundwater. *J. Hydrol.* **50**, 201–216 (1981).
- Ballentine, C. J. & Hall, C. M. An inverse technique for calculating paleotemperatures and other variables using noble gas concentrations in groundwater. *Geochim. Cosmochim. Acta* **63**, 2315–2336 (1999).
- Christiansen, J. E. Effect of entrapped air upon the permeability of soils. *Soil Sci.* **58**, 355–365 (1944).
- Fayer, M. J. & Hillel, D. Air encapsulation: 1. Measurement in a field soil. *Soil Sci. Soc. Am. J.* **50**, 568–572 (1986).
- Faybishenko, B. A. Hydraulic behavior of quasi-saturated soils in the presence of entrapped air: Laboratory experiments. *Wat. Resour. Res.* **31**, 2421–2435 (1995).
- Lehmann, B. E. & Pertschert, R. Radioisotope dynamics – the origin and fate of nuclides in groundwater. *Appl. Geochem.* **12**, 727–738 (1997).
- Schlosser, P., Stute, M., Sonntag, C. & Münnich, K. O. Tritogenic ^3He in shallow groundwater. *Earth Planet. Sci. Lett.* **94**, 245–256 (1989).
- Aeschbach-Hertig, W. *et al.* A $^3\text{H}/^3\text{He}$ study of ground water flow in a fractured bedrock aquifer. *Ground Wat.* **36**, 661–670 (1998).
- Dunkle Shapiro, S., Rowe, G., Schlosser, P., Ludin, A. & Stute, M. Tritium-helium 3 dating under complex conditions in hydraulically stressed areas of a buried-valley aquifer. *Wat. Resour. Res.* **34**, 1165–1180 (1998).
- Beyerle, U. *et al.* Infiltration of river water to a shallow aquifer investigated with $^3\text{H}/^3\text{He}$, noble gases and CFCs. *J. Hydrol.* **220**, 169–185 (1999).
- Solomon, D. K., Hunt, A. & Poreda, R. J. Source of radiogenic helium 4 in shallow aquifers: Implications for dating young groundwater. *Wat. Resour. Res.* **32**, 1805–1813 (1996).
- Osenbrück, K., Lippmann, J. & Sonntag, C. Dating very old pore waters in impermeable rocks by noble gas isotopes. *Geochim. Cosmochim. Acta* **62**, 3041–3045 (1998).
- Heaton, T. H. E., Talma, A. S. & Vogel, J. C. Origin and history of nitrate in confined groundwater in the western Kalahari. *J. Hydrol.* **62**, 243–262 (1983).
- Wilson, G. B. & McNeill, G. W. Noble gas recharge temperatures and the excess air component. *Appl. Geochem.* **12**, 747–762 (1997).
- Stute, M. & Talma, A. S. in *Isotope Techniques in the Study of Environmental Change 307–318* (IAEA, Vienna, Austria, 1998).
- Aeschbach-Hertig, W., Peeters, F., Beyerle, U. & Kipfer, R. Interpretation of dissolved atmospheric noble gases in natural waters. *Wat. Resour. Res.* **35**, 2779–2792 (1999).
- Aeschbach-Hertig, W., Stute, M., Schlosser, P., Clark, J. & Reuter, R. Large (9°C) glacial–interglacial temperature difference derived from an aquifer in Maryland (abstr.). *Eos* **77**, (Suppl.) S157 (1996).
- Mamyrin, B. A. & Tolstikhin, I. N. *Helium Isotopes in Nature* (Elsevier, Amsterdam, 1984).
- CLIMAP. The surface of the ice-age Earth. *Science* **191**, 1131–1137 (1976).

27. Rostek, F. *et al.* Reconstructing sea surface temperature and salinity using $\delta^{18}\text{O}$ and alkenone records. *Nature* **364**, 319–321 (1993).
28. Rind, D. & Peteet, D. Terrestrial conditions at the last glacial maximum and CLIMAP sea-surface temperature estimates: Are they consistent? *Quat. Res.* **24**, 1–22 (1985).
29. Thompson, L. G. *et al.* Late glacial stage and holocene tropical ice core records from Huascarán, Peru. *Science* **269**, 46–50 (1995).
30. Benson, B. B. & Krause, D. Isotopic fractionation of helium during solution: A probe for the liquid state. *J. Solut. Chem.* **9**, 895–909 (1980).

Acknowledgements

We thank J. Holocher and H. Baur for help in the laboratory and for discussions, and M. Stute, R. Reuter, K. Walraevens, J. Lermytte and C. Weyhenmeyer for collaboration in the field studies.

Correspondence and requests for materials should be addressed to W.A. (e-mail: aeschbach@eawag.ch).

The plastic deformation of iron at pressures of the Earth's inner core

H.-R. Wenk*, S. Matthies*, R. J. Hemley†, H.-K. Mao† & J. Shu†

* Department of Geology and Geophysics, University of California, Berkeley, California 94720, USA

† Geophysical Laboratory and Center for High-Pressure Research, Carnegie Institution of Washington, Washington DC 20015, USA

Soon after the discovery of seismic anisotropy in the Earth's inner core¹, it was suggested that crystal alignment attained during deformation might be responsible². Since then, several other mechanisms have been proposed to account for the observed anisotropy^{3,4}, but the lack of deformation experiments performed at the extreme pressure conditions corresponding to the solid inner core has limited our ability to determine which deformation mechanism applies to this region of the Earth⁵. Here we determine directly the elastic and plastic deformation mechanism of iron at pressures of the Earth's core, from synchrotron X-ray diffraction measurements of iron, under imposed axial stress, in diamond-anvil cells. The ϵ -iron (hexagonally close packed) crystals display strong preferred orientation, with c -axes parallel to the axis of the diamond-anvil cell. Polycrystal plasticity theory predicts an alignment of c -axes parallel to the compression direction as a result of basal slip, if basal slip is either the primary or a secondary slip system. The experiments provide direct observations of deformation mechanisms that occur in the Earth's inner core, and introduce a method for investigating, within the laboratory, the rheology of materials at extreme pressures.

Much of the solid Earth undergoes intense ductile deformation. Experiments at moderate pressure and temperature have provided information about deformation mechanisms of the important rock-forming minerals in the crust and upper mantle, with olivine, quartz and calcite receiving most attention. It has been well established that intracrystalline deformation causes reorientations of crystals. The orientation pattern (or texture) in a polycrystal is indicative of both the deformation history and the mechanisms. On a macroscopic scale, texture causes anisotropy of physical properties such as propagation of seismic waves.

Here we present a method for performing deformation experiments with diamond-anvil cells, analysing the polycrystal texture *in situ*, and interpreting from it the deformation mechanisms at pressures up to that of the Earth's core. It is generally agreed that the solid inner core is composed of a nickel-containing iron alloy. Theoretical^{6,7} as well as experimental studies^{8–11} have examined the phase diagram of iron at high pressure, and it is likely that iron exists

as ϵ -iron⁴. Application of this technique to iron at pressures above 200 GPa reveals ductile deformation mechanisms that may be the key to a better understanding of the origin of elastic anisotropy of the inner core.

In recent experiments, the conventional megabar diamond-anvil cell configuration was modified to observe diffraction from lattice planes oriented at different angles relative to the diamond-anvil axis^{12–14} (Fig. 1). The polychromatic synchrotron X-rays are incident on the sample at 84° to the diamond-anvil axis and diffraction patterns are recorded with an energy-dispersive detector in a symmetrical position ($2\theta = 12^\circ$). During an experiment, the cell is rotated around an axis perpendicular to the diamond-anvil axis to bring different lattice planes into diffraction condition. Two experiments were performed at the superconducting wiggler beamline X17C at Brookhaven National Laboratory. Run 1 reached 54 GPa; the 10- μm thick, 20- μm diameter iron powder sample, of grain-size 0.1–0.5 μm , was compressed in a beryllium gasket between flat diamond anvils of 300- μm diameter and probed by a 20- μm X-ray beam. Run 2 reached 220 GPa; the 5- μm thick by 15- μm diameter specimen was compressed between bevelled diamond anvils (500- μm outer diameter, 90- μm inner diameter, 9.5° bevel angle) and probed by a 15- μm X-ray beam. Each pattern took 10 minutes to record for the larger sample in run 1, and 60 minutes for the smaller sample in run 2. Figure 2 reveals regular variations in peak intensities with rotation angle χ . There are also systematic differences in d -spacings of spectra for lattice planes oriented perpendicular and parallel to the diamond axis. Shifts in d -spacings can be explained in terms of elastic deformation under stress¹⁴, whereas the variations in intensities are due to crystal alignment attained during plastic deformation. Whereas the intensity variations are obstacles for crystal structure refinements, they offer unique opportunities for interpreting mechanical characteristics.

For eight diffraction peaks (100, 002, 101, 102, 110, 200, 112 and 201) intensities were extracted for 10 sample orientations, corresponding to lattice planes oriented from parallel to perpendicular to the diamond-anvil axis in 10° intervals (angle χ). Intensities were integrated over the peak width in energy, and the background was subtracted. Because of incident beam intensity fluctuations, and differences in irradiated volume and absorption characteristics for different sample orientations, a preliminary normalization was performed by dividing each peak intensity by the summation of all peak intensities in a spectrum. From the relative intensity variations, the orientation distribution was calculated¹⁵. In the case of axial texture symmetry, results are most conveniently represented in inverse pole figures for the direction of the

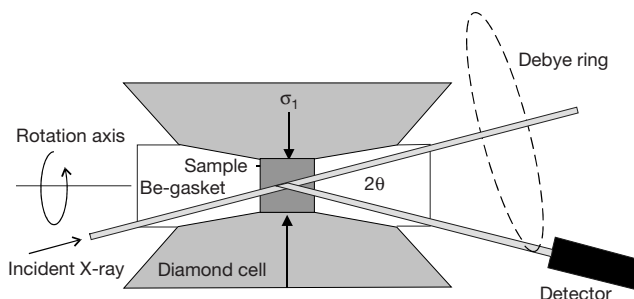


Figure 1 Experimental set-up. The deformation experiments are performed with a diamond-anvil cell positioned in a synchrotron X-ray beam. An axial stress (σ_1) that produces ductile deformation is imposed parallel to the diamond axis. The incident beam of polychromatic synchrotron X-rays and the energy-dispersive detector are arranged symmetrically with a 2θ diffraction angle of 12° . The diamond-anvil assembly is rotated around an axis perpendicular to the diamond-cell axis to record intensity variations along Debye diffraction rings.

diamond-cell axis (Fig. 3a, b). These diagrams document strong preferred orientations with a dominating (0001) fibre component, that is, *c*-axes of the hexagonal crystals are aligned parallel to the compression direction.

From the texture pattern of a deformed material, it is possible to infer active deformation systems by comparing experiments with corresponding polycrystal plasticity simulations¹⁶. In hexagonal metals, several slip and twinning systems have been identified: among them, (0001){1210} basal slip and {1010}{1210} prismatic slip are most important. Their relative activity depends on temperature, strain rate, and particularly on grain orientation. Hexagonal metals have been divided into two groups, those with a high *c/a* ratio such as Be, Zn and Cd and others with a low *c/a* ratio such as Zr and Ti. The former are elastically more anisotropic and generally basal slip is the primary slip system. Zr and Ti are elastically fairly isotropic and prismatic slip is the preferred slip system. Hexagonal ϵ -iron has a low *c/a* ratio (1.604 at 49 GPa)¹⁷ and therefore it may behave similarly to Zr and Ti. If the stacking fault energy rather than the *c/a* ratio determines the primary slip system, ϵ -iron should deform similarly to zinc⁵. We note that basal and prismatic slip have been observed in all hexagonal metals^{18–20}. In order to get some indication about the sensitivity of slip system activity on texture, we investigated the influence of different critical resolved shear stress ratios for basal and prismatic slip (Table 1).

Compared with cubic metals, hexagonal metals have a much higher plastic anisotropy, that is, different orientations deform with different ease. This has been a limitation when simulating texture development with compatibility-based models such as that of Taylor. Finite element approaches²¹ and self-consistent models²² have been successful in predicting texture development for Zr and Ti during rolling at ambient conditions. We have used the viscoplastic self-consistent theory in the anisotropic one-site version²² to simulate texture development of ϵ -iron in compression by imposing a uniform strain path (constant strain rate and no strain hardening)

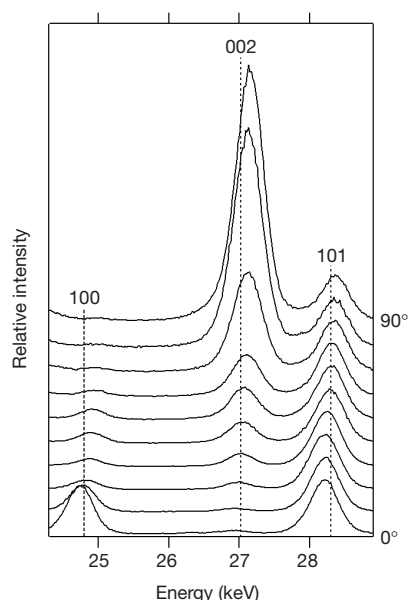


Figure 2 X-ray diffraction spectra of ϵ -iron recorded at 220 GPa and 300 K for different rotation angles, χ . Values of the rotation angle (from 0° to 90°) are indicated on the right-hand vertical axis. Intensity is plotted as a function of X-ray energy (in keV). At high χ (lattice planes perpendicular to the diamond cell axis), high intensities are observed for (002), and low intensities for (100); at low χ the opposite is true. Also notice systematic shifts in *d*-spacings towards higher energy (lower *d*-spacings) as the compression direction ($\chi = 90^\circ$) is approached.

to approximate deformation in the diamond-anvil cell. The simplified assumption is made that each grain orientation deforms as a viscoplastic inclusion in a homogeneous anisotropic medium that represents an average over the grain aggregate. The solution satisfies stress equilibrium and strain compatibility on an average scale, though not locally for each grain. Differently oriented grains deform on different slip systems and at different rates.

Inverse pole figures (Fig. 3c–h) illustrate that a fibre texture with a concentration of compression axes near $c = [0001]$ develops with increasing strain, no matter whether basal slip is the primary or secondary slip system. The plastic spin due to basal slip reorients the basal poles towards the most compressed direction. The spin due to prismatic slip causes a rotation around the (0001) direction, leaving the basal pole orientation invariant. If basal slip is favoured over prismatic slip, it speeds up the alignment of the basal poles with the

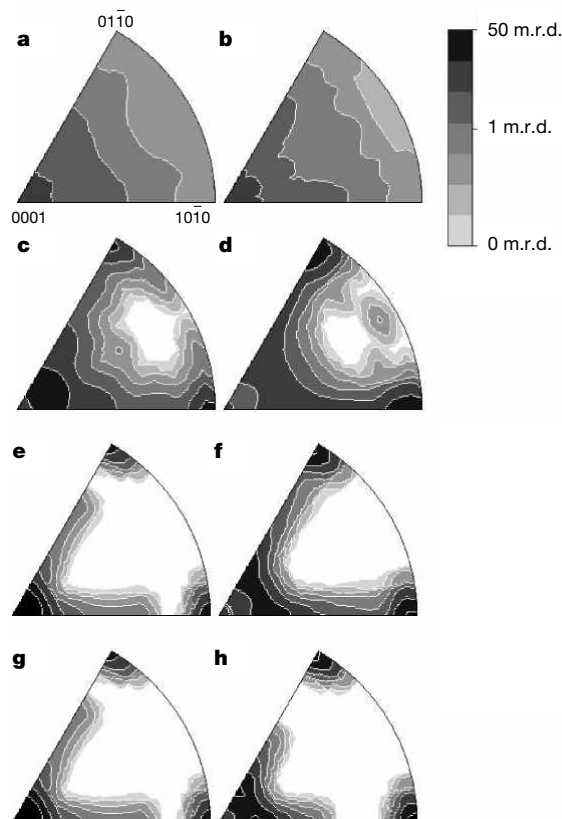


Figure 3 Inverse pole figures illustrating preferred orientation patterns of ϵ -iron deformed in axial compression. **a, b**, Experimental inverse pole figures of the diamond cell axis of ϵ -iron deformed at 54 GPa and 220 GPa. The maximum pole density for the 54-GPa sample is 5.78 multiples of a random distribution (m.r.d.), that for the 220-GPa sample is 7.32 m.r.d. **c–h**, Simulations of texture development, assuming different critical shear stress ratios for slip systems (prismatic–basal) as follows: 2–1 (**c, e, g**) and 0.5–1 (**d, f, h**) (see Table 1). Simulations were done with the self-consistent polycrystal plasticity model²² that requires as input data single-crystal slip systems, critical shear stress ratios, the initial grain orientations (in our case a random distribution), the strain rate sensitivity of the stress (stress exponent assumed to be 15), and a uniform deformation history defined as incremental displacement gradients. Strain steps of 50% (**c, d**), 100% (**e, f**) and 150% (**g, h**) are shown. In the model, twinning becomes activated at very large strains when other modes of deformation are not possible. When mechanical twinning does occur, the texture changes radically (not shown). In the inverse pole figures of the compression direction shown here, pole densities were obtained by smoothing the orientation distribution of 500 individual orientations with a 7.5° gaussian and normalizing the distribution. Equal-area projection is used, and logarithmic contours are expressed in m.r.d.

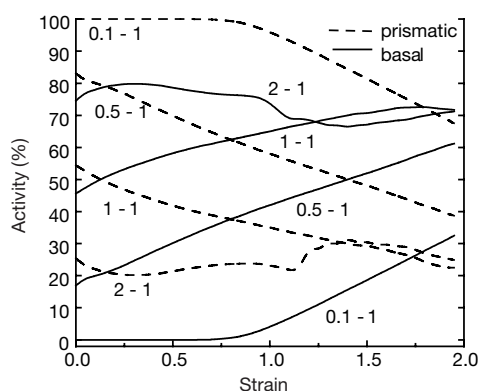


Figure 4 Calculated activities of basal and prismatic slip in ϵ -iron as a function of compressive strain. In these polycrystal plasticity simulations, we have used models with critical shear stress ratios as follows (prismatic–basal): 2–1, 1–1, 0.5–1 and 0.1–1 (Table 1). We note that for all models the activity of basal slip increases with large strain and that of prismatic slip decreases.

compressive direction but does not qualitatively change the texture pattern.

The results in Fig. 3c–h are in good agreement with the texture patterns observed in the diamond-anvil experiments (Fig. 3a, b). We conclude that a large plastic strain (50–100%) was attained, and that basal slip is an active deformation mechanism in ϵ -iron (in agreement with predictions⁵), probably in combination with other systems. Even if prismatic slip is favoured, basal slip becomes dominant as preferred orientation develops, as is illustrated in a graph of slip system activity (Fig. 4).

Many seismological studies have confirmed that compressional waves travel 3–4% faster along the (vertical) axis of the Earth than in the equatorial plane^{23–26}. Therefore it is of interest to estimate the seismic anisotropy in experimentally deformed ϵ -iron. We have calculated elastic properties for the 220-MPa polycrystalline ϵ -iron sample as a geometrical mean average of the elastic constants of the single crystal over the texture, assuming an elastically fairly isotropic single crystal⁷, and a much more anisotropic single crystal¹⁴. The P-wave anisotropies of 4% and 18%, for low and high single-crystal anisotropy, respectively, are within the range of anisotropies documented for the inner core: this suggests that ϵ -iron, deforming dominantly by basal slip, may be the cause. However, the large difference also highlights the need for the determination of more reliable single-crystal elastic constants at high pressure and temperature in order to understand core anisotropy. Such information could be obtained from a detailed analysis of *d*-spacing variations in stressed and textured polycrystals^{27,28}.

Iron in the inner core does not deform in a purely compressive geometry, but having identified deformation systems, it is possible to model much more complex and heterogeneous situations. For example, if deformation in the Earth's core occurs during convection and crystals deform by basal and prismatic slip, very strong

textures develop²⁹. Similar patterns are expected if deformation is caused by magnetic forces³⁰. So far such inferences have been largely hypothetical, and models are greatly aided by the availability of experimental data.

The present results represent the first step towards quantitative texture analysis at extreme pressure conditions. Extension of the method to simultaneous high-pressure/high-temperature conditions, as well as quantification of stress and strain, is feasible, and would provide a tool to study the rheology of rocks in the deep interior of the Earth at prevailing conditions. Results from such experiments and refined modelling would allow us to investigate not only slip, but also thermally activated processes such as recovery and recrystallization; they would also help in addressing the still unresolved question of whether pressure influences slip system and dislocation activity, and provide a test of theoretical predictions. □

Received 11 January; accepted 25 April 2000.

- Morelli, A., Dziewonski, A. M. & Woodhouse, J. H. Anisotropy of the core inferred from PKIKP travel times. *Geophys. Res. Lett.* **13**, 1545–1548 (1986).
- Jeanloz, R. & Wenk, H.-R. Convection and anisotropy of the inner core. *Geophys. Res. Lett.* **15**, 72–75 (1988).
- Jeanloz, R. The nature of the Earth's core. *Ann. Rev. Earth Planet. Sci.* **18**, 357–386 (1990).
- Stixrude, L. & Brown, J. M. in *Ultrahigh-Pressure Mineralogy: Physics and Chemistry of the Earth's Deep Interior* (ed. Hemley, R. J.) 261–282 (Reviews in Mineralogy, Vol. 37, Mineralogical Society of America, Washington DC, 1998).
- Poirier, J.-P. & Price, G. D. Primary slip system of ϵ -iron and anisotropy of the Earth's inner core. *Phys. Earth Planet. Inter.* **110**, 147–156 (1999).
- Soderlind, P., Moriarty, J. A. & Wills, J. M. First principles theory of iron up to earth-core pressures: structural, vibrational, and elastic properties. *Phys. Rev. B* **53**, 14063–14072 (1996).
- Stixrude, L. & Cohen, R. E. High-pressure elasticity of iron and anisotropy of Earth's inner core. *Science* **267**, 1972–1975 (1995).
- Saxena, S. *et al.* Synchrotron x-ray study of iron at high pressure and temperature. *Science* **269**, 1703–1704 (1995).
- Yoo, C. S., Akella, J., Campbell, A. J., Mao, H. K. & Hemley, R. J. Phase diagram of iron by in situ x-ray diffraction: implications for Earth's core. *Science* **270**, 1473–1475 (1995).
- Andrault, D., Fiquet, G., Kunz, M., Viscocekas, F. & Hausermann, D. The orthorhombic structure of iron: an in-situ study at high temperature and high pressure. *Science* **278**, 831–834 (1997).
- Shen, G. & Heinz, D. L. in *Ultrahigh-Pressure Mineralogy: Physics and Chemistry of the Earth's Deep Interior* (ed. Hemley, R. J.) 369–396 (Reviews in Mineralogy, Vol. 37, Mineralogical Society of America, Washington DC, 1998).
- Hemley, R. J. *et al.* X-ray imaging of stress and strain of diamond, iron, and tungsten at megabar pressures. *Science* **276**, 1242–1245 (1997).
- Singh, A. K., Mao, H. K., Shu, J. & Hemley, R. J. Estimation of single-crystal elastic moduli from polycrystalline x-ray diffraction at high pressure: application to FeO and iron. *Phys. Rev. Lett.* **80**, 2157–2160 (1998).
- Mao, H. K. *et al.* Elasticity and rheology of iron above 220 GPa and the nature of the Earth's inner core. *Nature* **396**, 741–743 (1998).
- Wenk, H.-R., Matthies, S., Donovan, J. & Chateigner, D. BEARTEX, a Windows-based program system for quantitative texture analysis. *J. Appl. Crystallogr.* **31**, 262–269 (1998).
- Kocks, U. F., Tomé, C. N. & Wenk, H.-R. *Preferred Orientations in Polycrystals and Their Effect on Materials Properties* (Cambridge Univ. Press, 1998).
- Mao, H. K., Wu, Y., Chen, L. C., Shu, J. F. & Jephcoat, A. P. Static compression of iron to 300 GPa and Fe₈Ni₁₂ alloy to 260 GPa: Implications for composition of the core. *J. Geophys. Res.* **95**, 21737–21742 (1990).
- Akhtar, A. Basal slip in zirconium. *Acta Metall.* **21**, 1–11 (1973).
- Greenspan, J. in *Beryllium, its Metallurgy and Properties* (ed. Hausner, H. H.) 240–246 (University of California Press, Berkeley, 1965).
- Tenckhoff, E. The development of the deformation texture in zirconium during rolling in sequential passes. *Metall. Trans. A* **9**, 1401–1412 (1978).
- Prantl, V. C., Jenkins, J. T. & Dawson, P. R. Modeling deformation induced textures in titanium, using analytical solutions for constrained single crystal response. *J. Mech. Phys. Solids* **43**, 1283–1302 (1995).
- Lebensohn, R. A. & Tomé, C. N. A self-consistent anisotropic approach for the simulation of plastic deformation and texture development of polycrystals – Application to zirconium alloys. *Acta Metall.* **41**, 2611–2624 (1993).
- Romanowicz, B., Li, X. D. & Durek, J. Anisotropy of the inner core: Could it be due to low-order convection? *Science* **274**, 963–966 (1996).
- Shearer, P. M. Constraints on inner core anisotropy from PKP(DF) travel times. *J. Geophys. Res.* **99**, 19647–19659 (1994).
- Song, X. Anisotropy of the Earth's inner core. *Rev. Geophys.* **35**, 297–313 (1997).
- Su, W.-J. & Dziewonski, A. M. Inner core anisotropy in three dimensions. *J. Geophys. Res.* **100**, 9831–9852 (1995).
- Hauk V. & Kockelmann, H. Evaluation of single crystal elastic constants from mechanical and X-ray elastic constants of the polycrystal. *Z. Metall.* **70**, 500–502 (1979).
- Bittorf, Ch., Matthies, S., Priesmeyer, H. G. & Wagner, R. Diffraction determination of thermo-elastic single crystal constants using polycrystalline samples. *Intermetallics* **7**, 1–8 (1998).
- Wenk, H.-R., Baumgardner, J. R., Lebensohn, R. A. & Tomé, C. N. A convection model to explain anisotropy of the inner core. *J. Geophys. Res.* **105**, 5663–5677 (2000).
- Karato, S. Seismic anisotropy of the Earth's inner core caused by the Maxwell stress-induced flow. *Nature* **402**, 871–873 (1999).

Table 1 Slip and twinning systems and their critical resolved shear stress ratios used in the texture simulations of ϵ -iron

	A	B	C	D
(0001)($\bar{1}2\bar{1}0$) slip	1	1	1	1
{10 $\bar{1}0$ }($\bar{1}2\bar{1}0$) slip	0.1	0.5	1	2
{10 $\bar{1}1$ }($\bar{1}2\bar{1}0$) slip	10	10	10	10
{10 $\bar{1}1$ }(11 $\bar{2}$ 3) slip	10	10	10	10
{21 $\bar{1}2$ }(211 $\bar{1}$ -3) C-twins	10	10	10	10
{10 $\bar{1}2$ }(10 $\bar{1}1$) T-twins	10	10	10	10

Here the values of critical resolved shear stress ratios are given assuming that the value of critical resolved shear stress for basal slip equals unity.

Acknowledgements

We thank A. K. Singh, S. Merkel and J. Hu for discussions and assistance. The work was supported by IGPP-LANL and NSF.

Correspondence and requests for materials should be addressed to H.-R.W. (e-mail: wenk@seismo.berkeley.edu).

Ecosystem size determines food-chain length in lakes

David M. Post^{*†}, Michael L. Pace[†] & Nelson G. Hairston Jr^{*}

^{*} Department of Ecology and Evolutionary Biology, Corson Hall, Cornell University, Ithaca, New York 14853, USA

[†] Institute of Ecosystem Studies, Box AB, Millbrook, New York 12545, USA

Food-chain length is an important characteristic of ecological communities¹: it influences community structure², ecosystem functions^{1–4} and contaminant concentrations in top predators^{5,6}. Since Elton⁷ first noted that food-chain length was variable among natural systems, ecologists have considered many explanatory hypotheses^{1,4,8,9}, but few are supported by empirical evidence^{4,10,11}. Here we test three hypotheses that predict food-chain length to be determined by productivity alone (productivity hypothesis)^{4,10,12,13}, ecosystem size alone (ecosystem-size hypothesis)^{14,15} or a combination of productivity and ecosystem size (productive-space hypothesis)^{7,16–18}. The productivity and productive-space hypotheses propose that food-chain length should increase with increasing resource availability; however, the productivity hypothesis does not include ecosystem size as a determinant of resource availability. The ecosystem-size hypothesis is based on the relationship between ecosystem size and species diversity, habitat availability and habitat heterogeneity^{14,15}. We find that food-chain length increases with ecosystem size, but that the length of the food chain is not related to productivity. Our results support the hypothesis that ecosystem size, and not resource availability, determines food-chain length in these natural ecosystems.

A major impediment to critically testing the productivity (Fig. 1a), productive-space (Fig. 1c) and ecosystem-size (Fig. 1b) hypotheses in natural systems has been the inability to measure accurately both food-chain length and ecosystem size^{12,18}. We overcame these obstacles by using stable isotope techniques to determine food-chain length and by taking advantage of the relative isolation of lakes to estimate ecosystem size. We estimated food-chain length in 25 northern temperate lakes ranging in volume from 3.8×10^5 to 1.7×10^{12} m³, and ranging in total phosphorus from 2.6 to 230 $\mu\text{g l}^{-1}$. Nearly all northern temperate lakes are phosphorus-limited and total phosphorus is a strong predictor of primary productivity in phosphorus-limited lakes¹⁹. Vander Zanden *et al.* described a positive relationship between food-chain length and both lake area and water clarity (their measure of productivity)²⁰. They could not separate, however, the influence of productivity and ecosystem size because their gradients of lake area and water clarity were correlated. Total phosphorus and lake volume are not correlated in our 25 lakes ($r = -0.21$, $P = 0.31$).

Our work differs from previous empirical tests of food-chain theory in four important ways. First, we collected data from lakes along independent gradients of productivity and ecosystem size. Second, our observations were made at the ecologically relevant scale of whole food webs. Third, our observations were made in temperate lakes that are ecologically similar, contain similar communities and are all located within a restricted geographic region.

Finally, we used stable isotope techniques to estimate maximum trophic position (MTP), a variable that is conceptually similar to mean food-chain length¹⁰. Because MTP is a continuous variable, we can detect subtle changes in food-chain length within the naturally occurring range of ecosystem size and productivity.

Stable isotope ratios of nitrogen and carbon are powerful tools for evaluating trophic structure and energy flow in ecological communities²¹. The $\delta^{15}\text{N}$ of an organism is typically enriched by 3.4‰ ($\pm 1\%$) relative to its diet²², and can be used to determine the trophic position of an organism. In contrast, $\delta^{13}\text{C}$ changes little as carbon moves through the food web and can be used to evaluate the ultimate sources of energy for an organism^{21,23}. In lakes, $\delta^{13}\text{C}$ is particularly useful for differentiating between the two major sources of available energy: littoral (near shore) production from attached algae and detritus, and pelagic (open water) production from phytoplankton²⁴. In our study lakes, the difference between littoral and pelagic $\delta^{13}\text{C}$ was between 2 and 10‰ (mean, 6.5%), with littoral $\delta^{13}\text{C}$ enriched in ^{13}C relative to pelagic $\delta^{13}\text{C}$.

Stable isotopes provide a continuous measure of trophic position (as opposed to discrete trophic levels) which integrates the assimilation of mass from all the trophic pathways leading to a top predator^{8,25}. Estimates of trophic position using stable isotope techniques therefore reflect the magnitude of energy or mass flow through different food web pathways and account for complex interactions such as trophic omnivory⁵. Trophic position is calculated as $\lambda + (\delta^{15}\text{N}_{\text{organism}} - \delta^{15}\text{N}_{\text{base of food web}})/3.4$, where λ is the trophic position of the organism used to estimate $\delta^{15}\text{N}_{\text{base of food web}}$ (for example, $\lambda = 2$ for primary consumers), $\delta^{15}\text{N}_{\text{organism}}$ is measured directly and 3.4 is the average enrichment in $\delta^{15}\text{N}$ per trophic level. For long-lived and mobile predators, $\delta^{15}\text{N}_{\text{base of food web}}$ (hereafter called $\delta^{15}\text{N}_{\text{base}}$) must capture the potentially high temporal variation in $\delta^{15}\text{N}$ of primary producers and detrital energy sources, and account for the spatial heterogeneity in $\delta^{15}\text{N}$ both within and among lakes. Our observations show that, within a lake, there can be seasonal differences of $>4\%$ in the $\delta^{15}\text{N}$ of phytoplankton and 3–4‰ differences in $\delta^{15}\text{N}_{\text{base}}$ between the littoral and pelagic food webs. Furthermore, we observed differences of $>13\%$ in $\delta^{15}\text{N}_{\text{base}}$ among lakes for both pelagic and littoral food webs. To account for this variability, we used filter-feeding mussels and surface-grazing

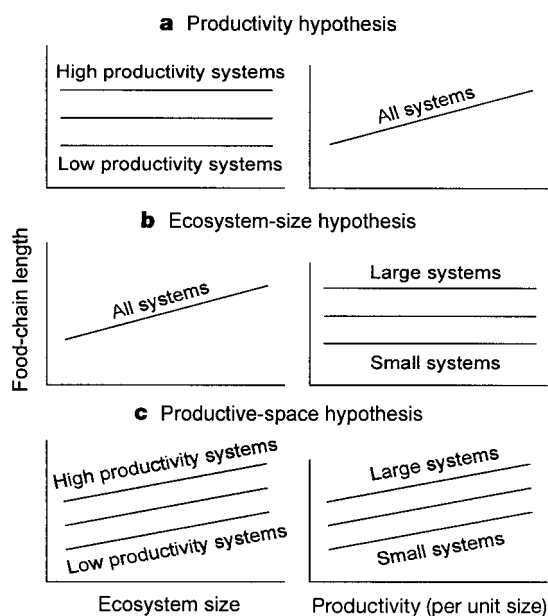


Figure 1 Hypothesized relationships between food-chain length and ecosystem size, and between food-chain length and productivity. **a**, For the productivity hypothesis; **b**, for the ecosystem-size hypothesis; **c**, for the productive-space hypothesis.

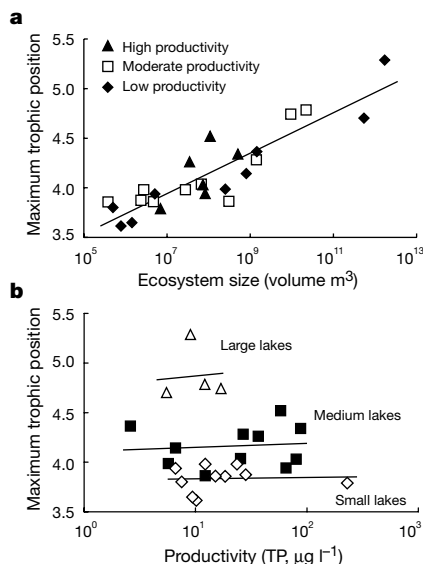


Figure 2 Relationships between maximum trophic position and ecosystem size or productivity. **a**, Ecosystem size for low ($2\text{--}11\ \mu\text{g l}^{-1}$ total phosphorus (TP)), moderate ($11\text{--}30\ \mu\text{g l}^{-1}$ TP) and high productivity lakes ($30\text{--}250\ \mu\text{g l}^{-1}$ TP). **b**, Productivity for small (3×10^5 to $3 \times 10^7\ \text{m}^3$), medium (3×10^7 to $3 \times 10^9\ \text{m}^3$) and large lakes (3×10^9 to $2 \times 10^{12}\ \text{m}^3$). Maximum trophic position is the trophic position of the species with the highest average trophic position in each of the lake food webs. The data are from 25 lakes in northeastern North America.

snails collected in each of our study lakes as temporal integrators of $\delta^{15}\text{N}_{\text{base}}$ for the pelagic and littoral food webs, respectively. Snails reflect the isotopic signature of the detritus and periphyton that form the base of littoral food webs and mussels reflect the isotopic signature of seston, which forms the base of pelagic food webs. We calculated the trophic position of fish at the top of each lake food web using the equation: trophic position = $2 + (\delta^{15}\text{N}_{\text{fish}} - [\delta^{15}\text{N}_{\text{mussel}} \times \alpha + \delta^{15}\text{N}_{\text{snail}} \times (1 - \alpha)]) / 3.4$, where α is the proportion of carbon in a target organism ultimately derived from the base of the pelagic food web: $\alpha = (\delta^{13}\text{C}_{\text{fish}} - \delta^{13}\text{C}_{\text{snail}}) / (\delta^{13}\text{C}_{\text{mussel}} - \delta^{13}\text{C}_{\text{snail}})$. Trophic position is an attribute of individual fish or of a single fish species, and MTP, like mean food-chain length, is an attribute of an entire food web. MTP is the trophic position of the species with the highest average trophic position in the lake food web.

Maximum trophic position increased with lake volume but not with total phosphorus (Fig. 2), and lake volume was the only significant predictor of MTP (log volume, $t = 9.28$, $P < 0.001$; log total phosphorus, $t = 0.59$, $P = 0.56$). Lake size alone explained 80% of the variation in maximum trophic position ($\text{MTP} = 2.51 + 0.2 \times \log \text{volume}$; $P < 0.001$, $r^2 = 0.80$). The increase in MTP from about 3.5 in the smallest lakes to around 5 in the Laurentian Great lakes represents an increase in food-chain length of almost 1.5 trophic levels. Four species occupied MTP in our data set: largemouth bass and northern pike in the smaller lakes, and walleye and lake trout in the largest lakes (Fig. 3a).

The increase in MTP with increasing lake size was caused by both the addition of new top predator species to larger lakes and a general increase in the trophic position of each top predator with increasing lake size (Fig. 3). For example, lake trout were not found in our smallest lakes, and in the largest lakes their trophic position increased with increasing lake size (Fig. 3). The addition of a new top predator increases MTP by adding new trophic steps to the top of the food web. In contrast, changes in the trophic position of a single top predator species must be caused by lengthening of food chains between the top and bottom of the food web. This lengthening is probably caused by some combination of functional

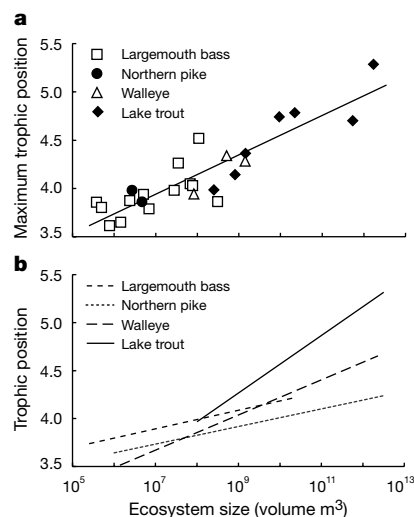


Figure 3 The increase in maximum trophic position is caused by both changes in top predator species and increases in the trophic position of each top predator. **a**, Maximum trophic position data from Fig. 2a labelled to identify the top predator species in each lake. **b**, The relationship (linear regression) between trophic position of top predator species and ecosystem size for largemouth bass (14 lakes, $n = 44$, $r^2 = 0.22$, $P < 0.01$), walleye (13 lakes, $n = 39$, $r^2 = 0.55$, $P < 0.01$), northern pike (11 lakes, $n = 25$, $r^2 = 0.20$, $P < 0.01$) and lake trout (7 lakes, $n = 22$, $r^2 = 0.89$, $P < 0.01$) plotted over the range of ecosystem size in which each species was found (data not shown).

diversification of the middle of the food web (for example, the addition of a new intermediate predator such as mysid shrimp, *Mysis relicta*, in our larger lakes) and reduced trophic omnivory at any or all trophic levels. Trophic omnivory probably declines as lake size increases because habitat heterogeneity and prey refugia increase with lake size. These changes may allow larger populations of preferred or optimal prey, which promote increased dietary specialization and reduced trophic omnivory. These changes to the top and middle of food webs are facilitated by increases in functional rather than species diversity.

If functional diversity is a determinant of maximum trophic position, then the ultimate cause of variation in food-chain length will be factors that influence the functional diversity of species in food webs. Human-mediated modifications of natural systems that affect the functional diversity of food webs, such as cultural eutrophication, species invasions and habitat fragmentation, will also affect food-chain length. Changes in food-chain length may, in turn, cause further changes in community dynamics, ecosystem function and contaminant concentrations in apical predators.

The importance of primary productivity in explaining variation in food-chain length has been debated for 40 years^{4,10,12,16}. Unlike the few studies that indicate some relationship between productivity and food-chain length^{12,20,26,27}, we found that ecosystem size was the only determinant of food-chain length in our study lakes. In contrast to many previous studies^{12,20,26,27}, we adopted a continuous measure of food-chain length and used independent gradients of productivity and ecosystem size in a natural setting. Although productivity must be an ultimate constraint of food-chain length^{16,17}, and may be important in small systems^{12,26} or where productivity is very low^{16,18}, our results and those of others^{4,8,10} show that productivity is a poor predictor of, and has a limited direct role in controlling food-chain length. □

Methods

To test the use of mussels and snails as indicators of the isotopic signature of the base of the littoral and pelagic food webs, we collected time series of periphyton (attached algae) and detritus samples from the littoral zone, and zooplankton and seston (a mixture of phytoplankton and other particulate organic matter) from integrated epilimnetic water

samples from the pelagic zones of three lakes (Spencer, Oneida and Cayuga). Periphyton and detritus were brushed from rocks, macrophytes and logs and pre-filtered through a 75- μm mesh to remove large invertebrates. Seston was pre-filtered through a 75 or 30 μm mesh. All periphyton and seston samples were then filtered onto pre-combusted glass fibre filters. Zooplankton were filtered from the water using a 150- μm mesh and visually inspected to remove particulate contaminants and predatory zooplankton. Each lake was sampled every two weeks from early June to late August (5–6 dates) in 1997 and 1998. Zooplankton were collected to provide the $\delta^{13}\text{C}_{\text{base}}$ of the pelagic food web because zooplankton are a better indicator of $\delta^{13}\text{C}_{\text{base}}$ for the pelagic food web than seston samples²⁸. Snails and mussels were sampled in late August. In Spencer Lake we used unionid mussels (Unionacea) and in Cayuga and Oneida lakes we used zebra mussels (*Dreissena polymorpha*). In three other New York lakes (Champlain, Conesus and Keuka lakes), where unionid and zebra mussels occurred together, we found no difference in their $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values (nested ANOVA with species nested in lake, d.f. = 3, $F = 2.85$, $P = 0.06$ for $\delta^{13}\text{C}$; d.f. = 3, $F = 1.79$, $P = 0.19$ for $\delta^{15}\text{N}$). Mussels and snails effectively captured the spatial variation and integrated the temporal variation in the $\delta^{13}\text{N}_{\text{base}}$ and $\delta^{13}\text{C}_{\text{base}}$ of pelagic and littoral food webs. Using lake-by-habitat combinations as replicates ($n = 6$ for both $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$), we found no significant differences between the median $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ of each time series and the $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ of snails and mussels (paired t -test for means: $t = 2.29$, $P = 0.07$ for $\delta^{13}\text{C}$; $t = 2.19$, $P = 0.08$ for $\delta^{15}\text{N}$, where we subtracted 3.4‰ from the $\delta^{15}\text{N}$ of snails and mussels to remove the expected one trophic level of enrichment).

In each lake, we collected all fish species that were likely to feed at the top of the food web. Because trophic position can increase with fish length, we collected adult fish of each species and held length as constant as possible across the lake size gradient. The fish species collected and the lengths of fish analysed were: largemouth bass (*Micropterus salmoides*; 250–440 mm), smallmouth bass (*Micropterus dolomieu*; 250–450 mm), northern pike (*Esox lucius*; 450–840 mm), chain pickerel (*Esox niger*; 390–530 mm), walleye (*Stizostedion vitreum*; 300–700 mm), burbot (*Lota lota*; 580–740 mm), lake trout (*Salvelinus namaycush*; 460–730 mm), brook trout (*Salvelinus fontinalis*; 410 mm), chinook salmon (*Oncorhynchus tshawytscha*; 800–1000 mm), rainbow trout (*Oncorhynchus mykiss*; 340–480 mm), Atlantic salmon (*Salmo salar*; 490–770 mm) and brown trout (*Salmo trutta*; 310–610 mm). Snails and mussels were collected from each lake between late July and early September each year. Most fish were collected in late July to October, but for a few lakes we used fish collected in June. We used total phosphorus from integrated epilimnetic water samples taken in late July or August as an index of lake productivity. Our range of total phosphorus (2.6 to 230 $\mu\text{g l}^{-1}$) corresponds to a range of primary productivity of ~30–450 $\text{g C m}^{-2} \text{ yr}^{-1}$ (ref. 29). We used previously documented volume estimates for 21 of our lakes. For the remaining four lakes, we estimated lake volume as a hyperbolic sinusoid: $0.43 \times \text{area} \times \text{maximum depth}$.

We took a small section of muscle tissue from each fish for isotopic analysis. Snails and mussels were dissected and aggregated, particulate contaminants were removed and only soft tissue was used for isotopic analysis. Samples were dried at 40 °C for >48 h and ground to a fine powder. We then extracted lipids (using methanol-chloroform extraction) from all animal samples because lipids are depleted in ^{13}C compared with whole organisms^{21,30} and lipid content in our tissue samples was variable (ranging from about 5% by mass in largemouth bass to >30% in some lake trout). Stable isotope analysis was performed using a Europa Geo 20/20 continuous flow isotope ratio mass spectrometer at the Cornell Laboratory for Stable Isotope Analysis. The standard error of the replicates of all our analyses were 0.05‰ for $\delta^{13}\text{C}$ and 0.18‰ for $\delta^{15}\text{N}$. All stable-isotope values are reported in the δ notation: $\delta^{15}\text{N} = ((^{15}\text{N}_{\text{sample}}/^{14}\text{N}_{\text{sample}})/(^{15}\text{N}_{\text{standard}}/^{14}\text{N}_{\text{standard}})) - 1 \times 1,000$, where the global standard is atmospheric nitrogen, and $\delta^{13}\text{C} = ((^{13}\text{C}_{\text{sample}}/^{12}\text{C}_{\text{sample}})/(^{13}\text{C}_{\text{standard}}/^{12}\text{C}_{\text{standard}})) - 1 \times 1,000$, where the global standard is Pee Dee Belemnite³¹.

Received 31 January; accepted 26 April 2000.

- Pimm, S. L. & Lawton, J. H. The number of trophic levels in ecological communities. *Nature* **275**, 542–544 (1977).
- Carpenter, S. R. & Kitchell, J. F. *The Trophic Cascade in Lakes* (Cambridge Univ. Press, Cambridge, 1993).
- DeAngelis, D. L. *Dynamics of Nutrient Cycling and Food Webs* (Chapman and Hall, London, 1992).
- Pimm, S. L. *Food Webs* (Chapman and Hall, London, 1982).
- Cabana, G. & Rasmussen, J. B. Modelling food chain structure and contaminant bioaccumulation using stable nitrogen isotopes. *Nature* **372**, 255–257 (1994).
- Kidd, K. A., Schindler, D. W., Hesslein, R. H. & Muir, D. C. G. Effects of trophic position and lipid on organochlorine concentrations in fishes from subarctic lakes in Yukon Territory. *Can. J. Fish. Aquat. Sci.* **55**, 869–881 (1998).
- Elton, C. *Animal Ecology* (Sidgwick and Jackson, London, 1927).
- Hairton, N. G. Jr & Hairton, N. G. Sr Cause–effect relationships in energy flow trophic structure and interspecific interactions. *Am. Nat.* **142**, 379–411 (1993).
- Persson, L., Bengtsson, J., Menge, B. A. & Power, M. A. in *Food Webs: Integration of Pattern and Process* (eds Polis, G. A. & Winemiller, K. O.) 396–434 (Chapman and Hall, New York, 1996).
- Briand, E. & Cohen, J. E. Environment correlates of food chain length. *Science* **238**, 956–960 (1987).
- Stern, R. W., Bajpai, A. & Adams, T. The enigma of food chain length: absence of theoretical evidence for dynamic constraints. *Ecology* **78**, 2258–2262 (1997).
- Kaunzinger, C. M. K. & Morin, P. J. Productivity controls food-chain properties in microbial communities. *Nature* **395**, 495–497 (1998).
- Oksanen, L., Fretwell, S. D., Arruda, J. & Niemelä, P. Exploitation ecosystems in gradients of primary productivity. *Am. Nat.* **118**, 240–261 (1981).
- Cohen, J. E. & Newman, C. M. Community area and food-chain length: theoretical predictions. *Am. Nat.* **138**, 1542–1554 (1992).
- Holt, R. D. in *Species Diversity in Ecological Communities* (eds Ricklefs, R. E. & Dchluter, D.) 77–88 (Univ. Chicago Press, Chicago, 1993).

- Hutchinson, G. E. Homage to Santa Rosalia; or, why are there so many kinds of animals? *Am. Nat.* **93**, 145–159 (1959).
- Slobodkin, L. B. *Growth and Regulation of Animal Populations* (Holt, Rinehart and Wilson, New York, 1961).
- Schoener, T. W. Food webs from the small to the large. *Ecology* **70**, 1559–1589 (1989).
- Schindler, D. W. Factors regulating phytoplankton production and standing crop in the world's lakes. *Limnol. Oceanogr.* **23**, 478–486 (1978).
- Vander Zanden, M. J., Shuter, B. J., Lester, N. & Rasmussen, J. B. Patterns of food chain length in lakes: a stable isotope study. *Am. Nat.* **154**, 406–416 (1999).
- Peterson, B. J. & Fry, B. Stable isotopes in ecosystem studies. *Ann. Rev. Ecol. Syst.* **18**, 293–320 (1987).
- Minagawa, M. & Wada, E. Stepwise enrichment of ^{15}N along food chains: further evidence and the relation between $\delta^{15}\text{N}$ and animal age. *Geochim. Cosmochim. Acta* **48**, 1135–1140 (1984).
- France, R. L. & Peters, R. H. Ecosystem differences in the trophic enrichment of ^{13}C in aquatic food webs. *Can. J. Fish. Aquat. Sci.* **54**, 1255–1258 (1997).
- France, R. L. Differentiation between littoral and pelagic food webs in lakes using carbon isotopes. *Limnol. Oceanogr.* **40**, 1310–1313 (1995).
- Paine, R. T. Food webs: road maps of interactions or grist for theoretical development? *Ecology* **69**, 1648–1654 (1988).
- Jenkins, B., Kitching, R. L. & Pimm, S. L. Productivity, disturbance and food web structure at a local spatial scale in experimental container habitats. *Oikos* **65**, 249–255 (1992).
- Persson, L., Diehl, S., Johansson, L., Andersson, G. & Hamrin, S. F. Trophic interactions in temperate lake ecosystems: a test of food chain theory. *Am. Nat.* **140**, 59–84 (1992).
- del Giorgio, P. A. & France, R. L. Ecosystem-specific patterns in the relationship between zooplankton and POM or microplankton $\delta^{13}\text{C}$. *Limnol. Oceanogr.* **41**, 359–365 (1996).
- Vollenweider, V. R. A. Das Nährstoffbelastungskonzept als Grundlage für den externen eingriff in den eutrophierungsprozeß stehender gewässer und talsperren. *Z. Wasser-u. Abwasser-Forschung* **12**, 46–56 (1979).
- Kling, G. W., Fry, B. & O'Brien, W. J. Stable isotopes and planktonic trophic structure in arctic lakes. *Ecology* **73**, 561–566 (1992).

Acknowledgements

We thank the New York State Department of Environmental Conservation fisheries biologists and managers, as well as J. Beemer, T. Butler, E. Mills, L. Rudstam, D. Parish, B. Pientka, L. Puth, M. Vogelsang and N. Vochick for their help collecting fish and other samples; K. L. Lovell, J. Burdet and C. Alpha for laboratory and field assistance; and T. E. Dawson, L. O. Hedin, M. H. Olson and L. M. Puth for comments. This research was supported by the Kieckhefer Adirondack Fellowship, the National Science Foundation Research Training Group for Biogeochemistry and Environmental Change at Cornell University and the National Science Foundation Graduate Research Training grant for Human Accelerated Environmental Change at Cornell University and the Institute of Ecosystem Studies.

Correspondence and requests for materials should be addressed to D.M.P. (e-mail: dmp18@cornell.edu).

Cryptophyte algae are robbed of their organelles by the marine ciliate *Mesodinium rubrum*

Daniel E. Gustafson Jr, Diane K. Stoecker, Matthew D. Johnson, William F. Van Heukelem & Kerri Sneider

University of Maryland, Center for Environmental Science, Horn Point Laboratory, PO Box 775, Cambridge, Maryland 21613, USA

Mesodinium rubrum (Lohmann 1908) Jankowski 1976 (= *Myrionecta rubra*)^{1,2} is a common photosynthetic marine planktonic ciliate which can form coastal red-tides³. It may represent a 'species complex'^{4,5} and since Darwin's voyage on the *Beagle*, it has been of great cytological, physiological and evolutionary interest⁴. It is considered to be functionally a phytoplankton because it was thought to have lost the capacity to feed and possesses a highly modified algal endosymbiont^{5,6}. Whether *M. rubrum* is the result of a permanent endosymbiosis or a transient association between a ciliate and an alga is controversial⁷. We conducted 'feeding' experiments to determine how exposure to a cryptophyte alga affects *M. rubrum*. Here we show that although *M. rubrum* lacks a cytostome (oral cavity)⁸, it ingests cryptophytes and steals their organelles, and may not maintain a permanent endosymbiont.

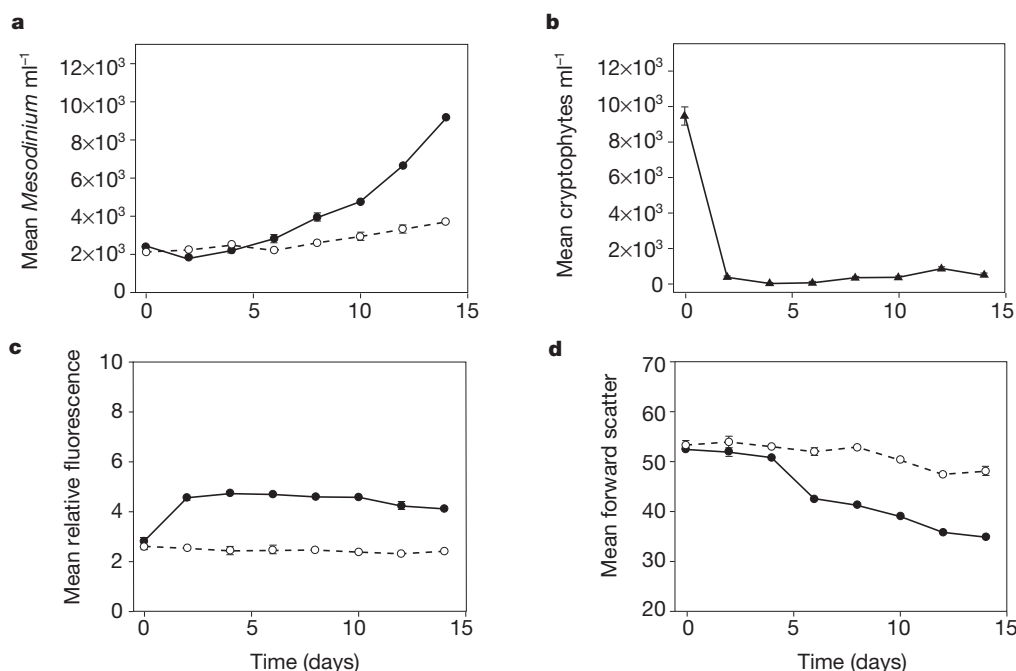


Figure 1 Changes in *Mesodinium rubrum* and cryptophyte populations measured by flow cytometry. Filled circles, cryptophyte-treated cultures; open circles, control cultures.

a, Abundance of *M. rubrum* cells. **b**, Abundance of cryptophyte cells. **c**, Mean relative

fluorescence of *M. rubrum*. **d**, Forward scatter of *M. rubrum*. (mean $\pm$ s.d., s.d. is not shown when it is no larger than the symbols).

M. rubrum does not fall into recognized cellular or functional categories, but may be a chimaera partially supported by organelle robbery.

M. rubrum contains unusual 'incomplete symbionts' consisting of numerous functional chloroplasts associated with non-ciliate mitochondria but apparently lacking nuclei⁹. Ultrastructural and pigment studies show that the chloroplasts are of cryptophycean origin¹⁰. The endosymbiont was thought to be highly modified, with little relationship to its free-living relations⁶. In contrast to the plastid-retaining ciliates^{11,12}, *M. rubrum* has a greatly reduced cytosome and no obvious digestive vacuoles⁸. It has never been observed to feed and there is no evidence of food particles inside the cell¹³. However, a related blue-pigmented *Mesodinium* sp. does capture cryptophytes¹⁴.

Dense blooms of *M. rubrum* often result in non-toxic red-tides and have been associated with extremely high rates of primary production in estuaries, fjords and upwelling areas^{5,15,16}. *M. rubrum* red-tides resulted in some of the highest recorded values for chlorophyll *a* and primary production in the marine environment¹⁷. There is good evidence for obligate phototrophy in *M. rubrum* from bloom studies, including high photosynthetic rates and uptake of inorganic nutrients^{3,18,19}. This photosynthetic ciliate is often an important primary producer in coastal and upwelling environments even when it does not form red-tides^{20–22}.

M. rubrum is very fragile and difficult to culture, so previous studies have used field assemblages^{23,24}. However, we obtained a culture of this ciliate from an enrichment of sea-ice/water collected in January 1996 from McMurdo Sound, Antarctica. We grew the isolate in algal growth media at 2–6 °C in the light. When grown in these conditions and supplied with a polar cryptophyte *Teleaulax acuta*, *M. rubrum* reaches densities of over 1.5×10^3 ml⁻¹, but it shows no sustained growth in the absence of algal 'prey'.

In the first experiment, we split a *M. rubrum* culture that had not been fed for 28 days into two; with (fed) and without (unfed, control) the addition of cryptophytes (10^4 ml⁻¹). The cultures were incubated for 14 days at 3 °C in 50–60 μ mol photons m⁻² s⁻¹ PAR (photosynthetically active radiation, ~400–700 nm). The fed culture exhibited a higher sustained growth rate (0.19 divisions

per day) than the control culture (0.09 divisions per day) (Fig. 1a). After addition of cryptophytes on day 1, *T. acuta* cells were reduced by 80% in 48 h (Fig. 1b). Red (chlorophyll *a*) and orange (phycoerythrin) fluorescence per *M. rubrum* cell increased by 1.5 times within 48 h and remained constant for 14 days (Fig. 1c). The forward scatter (an indicator of size) of *M. rubrum* cells decreased by 20% after 4–5 days since the cryptophyte addition and remained lower than in the control for the duration of the experiment (Fig. 1d).

Cultured *M. rubrum* cells are large (22–29 by 22–36 μ m) with volume 1,400–4,900 μ m³ cell⁻¹. At addition, the average bio-volumes were not significantly different between cultures (unfed 2,988 μ m³; fed 2,798 μ m³), whereas at day 6, there was a significant decrease (analysis of variance (ANOVA), $P < 0.05$) in the volume of the cryptophyte-treated cells (unfed 3,142 μ m³; fed 2,166 μ m³). By the end of the experiment, *M. rubrum* cells in the control culture were significantly larger (ANOVA, $P < 0.05$) than in the fed culture (2,716 μ m³ and 1,896 μ m³, respectively). Although *M. rubrum* can sustain a greater cell volume in the absence of cryptophytes, exposure to cryptophytes promotes cell division, resulting in a decrease in average cell size, but an increase in population size and biomass.

By day 6 there was a dramatic colour difference between the treatments; the fed culture was bright pink (similar to the cryptophyte) whereas the control culture was colourless. We examined 50 *M. rubrum* cells with epifluorescence microscopy at 0.5, 1.0, 4.0, 12, 24 h and 14 days after feeding. There were no fluorescence-blocking chlorophyll degradation bodies, evidence of digestion of algal 'prey', although these are observed in the digestive vacuoles of mixotrophic oligotrichous ciliates (D.K.S., personal observation).

In a second experiment, the percentage of *M. rubrum* with cryptophyte nuclei increased dramatically over the first hour after exposure to about 10^4 *T. acuta* ml⁻¹ after not being fed for 14 days (Fig. 2a). The *M. rubrum* cells lacked cryptophyte nuclei at the beginning of the incubation. Within 5 min, about 50% of the *M. rubrum* cells had one cryptophyte nucleus and by 60 min, about 20% of the ciliates had three or more cryptophyte nuclei (Fig. 2a). Within a few minutes of exposure to cryptophytes, 40% of

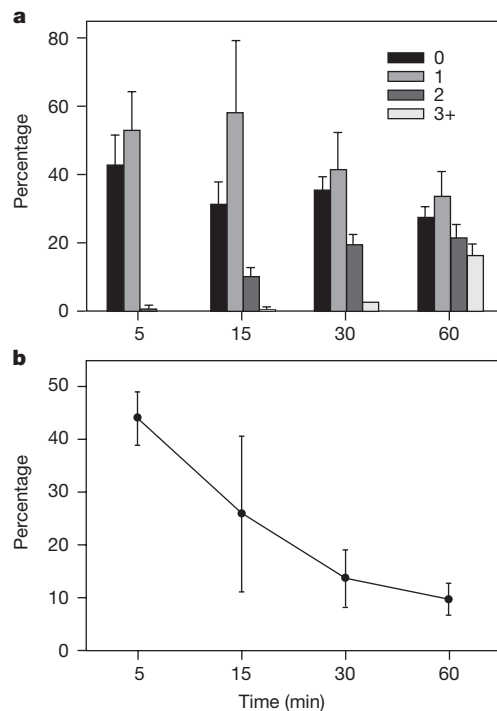


Figure 2 Ingestion of cryptophytes by *M. rubrum* when exposed to about 10^4 cryptophytes ml^{-1} after not being fed for 14 days. **a**, Percentage of *M. rubrum* containing cryptophyte nuclei. 0, no cryptophyte nuclei; 1, one cryptophyte nucleus; 2, two cryptophyte nuclei; 3+, three to seven cryptophyte nuclei present in one *M. rubrum* cell. **b**, Percentage of *M. rubrum* in the process of ingesting cryptophytes (mean $\pm$ s.d.).

the ciliates were in the process of ingesting cryptophytes, a percentage which decreased over time (Fig. 2b). During the first hour, the *M. rubrum* exhibited an ingestion rate of 1.3 ± 0.03 cryptophytes $\text{cell}^{-1} \text{h}^{-1}$ (mean $\pm$ s.d.) and a clearance of approximately $128 \text{ nl cell}^{-1} \text{h}^{-1}$.

In a third experiment, we added *T. acuta* to an *M. rubrum* culture that had not been fed for 14 days. Photosynthetic parameters were measured for fed and control cultures at day 10 and 17 of the incubation. Average chlorophyll content of *M. rubrum* was significantly higher in the fed culture than in the control on day 10 but not on day 17 (Fig. 3a). The *M. rubrum* chlorophyll *a* increased between day 10 and 17 from 223 to 394 ng ml^{-1} in the control and from 461 to 686 ng ml^{-1} in the fed culture. The average photosynthetic rate was not different between the fed and control cultures on day 10, but on day 17 the rate of photosynthesis per *M. rubrum* cell was greater in the fed culture than in the control (Fig. 3b). On day 17, the average chlorophyll-specific rates of photosynthesis of *M. rubrum* from the fed culture was significantly higher, $0.22 \pm 0.03 \text{ pg C (pg chlorophyll } a)^{-1} \text{ h}^{-1}$ than in the control, $0.12 \pm 0.03 \text{ pg C (pg chlorophyll } a)^{-1} \text{ h}^{-1}$ ($P < 0.05$).

Our results show that *M. rubrum* ingests free-living algae. Uptake of cryptophyte organelles is necessary for the sustained rapid growth of our isolate. Thus, the availability of cryptophyte 'prey' may trigger *M. rubrum* blooms and blooms may be partially limited by the availability of cryptophytes as a source of new organelles or nutrition. Whole cryptophytes may be present within the ciliate for a while after ingestion, but plastids are preferentially retained over cryptophyte nuclei. We did not observe chlorophyll degradation bodies, an indicator for digestion of algal prey, and digestive vacuoles have not been reported in investigations of this species with TEM (transmission electron microscopy)⁷⁻⁹. *M. rubrum*'s physiological ecology is very different from that of plastid-retaining ciliates, which are obligate mixotrophs and regularly need to ingest and digest prey to survive^{25,26}.

Increase in chlorophyll content in cultures deprived of prey for more than 16 days indicates that chlorophyll *a* can be synthesized

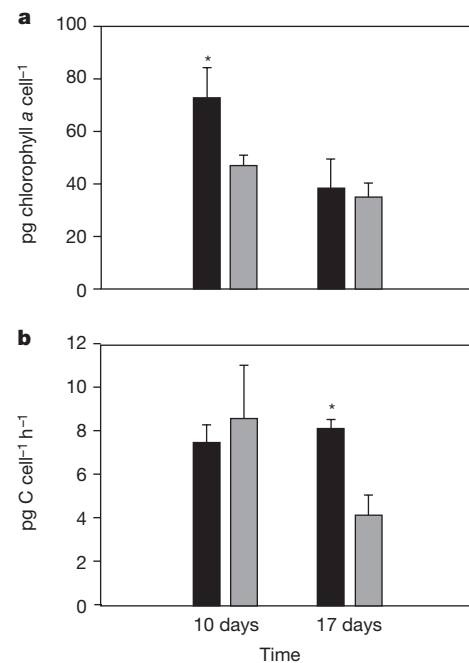


Figure 3 Chlorophyll content per cell and photosynthetic performance of *M. rubrum* at 10 and 17 days after addition of cryptophytes. Black bars, cryptophyte-treated culture; grey bars, control (untreated) culture. **a**, The chlorophyll *a* content. **b**, Photosynthetic performance per cell at $28\text{--}33 \mu\text{mol photons m}^{-2} \text{ s}^{-1}$ (mean $\pm$ s.d.). Asterisk, significant difference between the two cultures ($P < 0.05$).

within the ciliate. However, the uptake of cryptophyte plastids and the observed decrease in photosynthesis and growth of cells which had not recently been fed indicates that *M. rubrum* may need to feed periodically to replace ageing chloroplasts or chloroplasts diluted out by cell division. Although *M. rubrum* is photosynthetic and can synthesize chlorophyll *a*, it may not be an example of permanent endosymbiosis between an alga and a ciliate. It may be possible for the 'endosymbiont' to undergo degradation after ingestion by the ciliate and for the chloroplast and other organelles to persist for relatively long periods²⁷. Thus, *M. rubrum* may be a cell chimaera dependent on periodic ingestion of cryptophyte algae. From our data it is not possible to determine if endosymbiont or plastid reproduction occurs in *M. rubrum*. It is possible that when stolen organelles become non-functional, they are disposed of through egestion or digestion, although there is no direct evidence for either. In contrast to plastid-retaining ciliates, but similarly to algae, *M. rubrum* is able to use inorganic nutrients from the water column¹⁹. Thus, our isolate of *M. rubrum* does not fall into recognized categories for functional types of planktonic organisms.

Methods

Feeding experiments

We used triplicate incubation bottles for each treatment. We collected and preserved samples in 2% glutaraldehyde. We quantified cell numbers and cell attributes by using Coulter EPICS Profile II and Becton Dickinson FACSCalibur flow cytometers. For microscopy, we stained subsamples with the DNA specific stain, 4,6-diamidino-2-phenylindole (DAPI), and examined them with Zeiss blue 450–490 nm filter set for observation of plastids and degradation bodies and with Zeiss UV 365nm filter set for observation of nuclei.

¹⁴C techniques

We used triplicate incubation bottles for both the fed and control treatments. We added ¹⁴C-bicarbonate (final activity about $1.0 \mu\text{Ci ml}^{-1}$) to subsamples from each replicate. At the time of subsampling, there were $140 \text{ free cryptophytes ml}^{-1}$. After adding ¹⁴C, we took aliquots to determine total activity and split the replicates into dark (wrapped with

aluminum foil) and light bottles (28–33 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$) and incubated them for 24 h at 3 °C. We isolated individual ciliates from each light and dark replicate, washed them three times with sterile media (10 ml) and transferred ten washed cells into a scintillation vial. During the 2–3 h isolation period we maintained the samples in the dark on ice. We prepared the samples for liquid scintillation counting as described²⁸. We calculated rates of photosynthesis by subtracting average ¹⁴C fixation in the dark from fixation in the light. For the determination of *M. rubrum* chlorophyll, we isolated and washed ten cells before transferring them into vials containing cold 90% acetone and incubating them at –20 °C overnight for extraction. We measured chlorophyll *a* using a 10-AU Turner fluorometer.

Received 15 February; accepted 2 May 2000.

- Small, E. B. & Lynn, D. H. In *Illustrated Guide to the Protozoa* (eds Lee, J. J., Hunter, S. H. & Bovee, E. C.) 393–575 (Allen Press, Kansas, 1985).
- Krainer, K. -H. & Foissner, W. Revision of the genus *Askenasia* Blochmann, 1895, with proposal of two new species, and description of *Rhabdoaskenia minima* N. G., N. Sp. (Ciliophora, Cyclotrichida). *J. Protozool.* **37**, 414–427 (1990).
- Montagnes, D. J. S. & Lynn, D. H. The annual cycle of *Mesodinium rubrum* in the waters surrounding the Isles of Shoals, Gulf of Maine. *J. Plank. Res.* **11**, 193–201 (1989).
- Lindholm, T. in *Advances in Aquatic Microbiology*. Vol. 3 (eds Jannasch, H. W. & Williams, P. J.) 1–48 (Academic, London 1985).
- Crawford, D. W. *Mesodinium rubrum*: the phytoplankton that wasn't. *Mar. Ecol. Prog. Ser.* **58**, 161–174 (1989).
- Oakley, B. R. & Taylor, F. J. R. Evidence for a new type of endosymbiotic organization in a population of the ciliate *Mesodinium rubrum* from British Columbia. *BioSystems* **10**, 361–369 (1978).
- Hibberd, D. J. Observations on the ultrastructure of the cryptomonad endosymbiont of the red-water ciliate *Mesodinium rubrum*. *J. Mar. Biol. Assoc. UK* **57**, 45–61 (1977).
- Lindholm, T., Lindroos, P. & Mörk, A.-C. Ultrastructure of the photosynthetic ciliate *Mesodinium rubrum*. *BioSystems* **21**, 141–149 (1988).
- Taylor, F. J. R., Blackbourn, D. J. & Blackbourn, J. The red-water ciliate *Mesodinium rubrum* and its "incomplete symbionts": a review including new ultrastructural observations. *J. Fish. Res. Bd Can.* **28**, 391–407 (1971).
- Barber, R. T., White, A. W. & Siegelman, H. W. Evidence for a cryptomonad symbiont in the ciliate, *Cyclotrichum meunieri*. *J. Phycol.* **5**, 86–88 (1969).
- Laval-Peuto, M., Salvano, P., Gayol, P. & Gruet, C. Mixotrophy in marine planktonic ciliates: ultrastructure study of *Tontonia appendiculariformis* (Ciliophora, Oligotrichina). *Mar. Microb. Food Webs* **1**, 81–104 (1986).
- Stoecker, D. K. & Silver, M. W. Replacement and aging of chloroplasts in *Strombidium capitatum* (Ciliophora: Oligotrichida). *Mar. Biol.* **107**, 491–502 (1990).
- Lindholm, T. & Mörk, A.-C. Symbiotic algae and plastids in planktonic ciliates. *Mem. Soc. Fauna Flora Fennica* **65**, 17–22 (1989).
- Hargraves, P. Narrow River phytoplankton. *Maritimes* **35**, 6–8 (1991).
- Packard, T. T., Blasco, D. & Barber, R. T. in *Upwelling Ecosystems* (eds Boje, R. & Tomczak, M.) 73–89 (Springer, New York, 1978).
- Crawford, D. W., Purdie, D. A., Lockwood, A. P. M. & Weissman, P. Recurrent red-tides in the Southampton Water estuary caused by the phototrophic ciliate *Mesodinium rubrum*. *Est. Coast. Shelf Sci.* **45**, 799–812 (1997).
- Smith, W. O. & Barber, R. T. A carbon budget for the autotrophic ciliate *Mesodinium rubrum*. *J. Phycol.* **15**, 27–33 (1979).
- Wilkerson, F. P. & Grunseich, G. Formation of blooms by the symbiotic ciliate *Mesodinium rubrum*: the significance of nitrogen uptake. *J. Plank. Res.* **12**, 973–989 (1990).
- Kifle, D. & Purdie, D. A. The seasonal abundance of the phototrophic ciliate *Mesodinium rubrum* in Southampton Water, England. *J. Plank. Res.* **15**, 823–833 (1993).
- Stoecker, D. K., Taniguchi, A. & Michaels, A. E. Abundance of autotrophic, mixotrophic and heterotrophic planktonic ciliates in shelf and slope waters. *Mar. Ecol. Prog. Ser.* **50**, 241–254 (1989).
- Stoecker, D. K., Putt, M., Davis, L. H. & Michaels, A. E. Photosynthesis in *Mesodinium rubrum*: species-specific measurements and comparison to community rates. *Mar. Ecol. Prog. Ser.* **73**, 245–252 (1991).
- Sanders, R. W. Seasonal distributions of the photosynthesizing ciliates *Laboea strobila* and *Myrionecta rubra* (= *Mesodinium rubrum*) in an estuary of the Gulf of Maine. *Aquatic Microb. Ecol.* **9**, 237–242 (1995).
- Crawford, D. W. & Lindholm, T. Some observations on vertical distribution and migration of the phototrophic ciliate *Mesodinium rubrum* (= *Myrionecta rubra*) in a stratified brackish inlet. *Aquatic Microb. Ecol.* **13**, 267–274 (1997).
- Yih, W. & Shim, J. H. The planktonic phototrophic ciliate, *Mesodinium rubrum*, as a useful organism for marine biotechnological applications. *J. Mar. Biotechnol.* **5**, 82–85 (1997).
- Stoecker, D. K. in *Protozoa and Their Role in Marine Processes* Vol. G **25** (eds Reid, P. C. et al.) 161–179 (NATO ASI Series, Springer, Berlin–Heidelberg, 1991).
- Stoecker, D. K., Michaels, A. E. & Davis, L. H. Large proportion of marine planktonic ciliates found to contain functional chloroplasts. *Nature* **326**, 790–792 (1987).
- White, A. W., Sheath, R. G. & Hellebust, J. A. A red tide caused by the marine ciliate *Mesodinium rubrum* in Passamaquoddy Bay, including pigment and ultrastructure studies of the endosymbiont. *J. Fish. Res. Bd Can.* **34**, 413–416 (1977).
- Stoecker, D. K., Silver, M. W., Michaels, A. E. & Davis, L. H. Obligate mixotrophy in *Laboea strobila*, a ciliate which retains chloroplasts. *Mar. Biol.* **99**, 415–423 (1988).

Acknowledgements

We thank P. Del Giorgio for discussions; D. W. Coats, P. Krugens and G. Sellers for taxonomic assistance; and A. Li for isolating the cryptophyte. This work was partially supported by the US NSF Polar and Biological Oceanography (D.K.S.) and Maryland Sea Grant (K.S.).

Correspondence and requests for material should be addressed to D.E.G. (e-mail: gustafsn@hpl.umces.edu).

Language trees support the express-train sequence of Austronesian expansion

Russell D. Gray & Fiona M. Jordan

Department of Psychology, University of Auckland, Auckland 92019, New Zealand

Languages, like molecules, document evolutionary history. Darwin¹ observed that evolutionary change in languages greatly resembled the processes of biological evolution: inheritance from a common ancestor and convergent evolution operate in both. Despite many suggestions^{2–4}, few attempts have been made to apply the phylogenetic methods used in biology to linguistic data. Here we report a parsimony analysis of a large language data set. We use this analysis to test competing hypotheses—the “express-train”⁵ and the “entangled-bank”^{6,7} models—for the colonization of the Pacific by Austronesian-speaking peoples. The parsimony analysis of a matrix of 77 Austronesian languages with 5,185 lexical items produced a single most-parsimonious tree. The express-train model was converted into an ordered geographical character and mapped onto the language tree. We found that the topology of the language tree was highly compatible with the express-train model.

There are many parallels between the processes of biological and linguistic evolution and the methods used to analyse them⁴. Despite these parallels, however, historical linguists have not used the quantitative phylogenetic methods that have revolutionized evolutionary biology in the past 20 years⁸. So, although linguists routinely use the “comparative method”⁹ to construct language family trees from discrete lexical, morphological and phonological data, they do not use an explicit optimality criterion to select the best tree, nor do they typically use an efficient computer algorithm to search for the best tree from the discrete data. This is surprising given that the task of finding the best tree is inherently a combinatorial optimization problem of considerable computational difficulty¹⁰. One potential problem with a quantitative phylogenetic approach to linguistic evolution arises from the more reticulate nature of cultural evolution. Some authors^{11,12} have claimed that reticulate processes in linguistic evolution overshadow those of descent, leading them to reject the appropriateness of the family-tree model. We believe that this is an empirical claim, which can be evaluated using phylogenetic methods. If the data fit well on the tree and there is little systematic conflicting signal, then the family-tree model is supported. If the data fit poorly, then alternative phylogenetic methods that do not assume a tree model, such as spectral analysis or split decomposition, should be investigated. A critical part of phylogenetic inference involves testing for congruence between independent lines of evidence. Here we test a model of the colonization of the Pacific that is derived from predominantly archaeological data by quantitatively examining its fit with a parsimony tree of Austronesian languages.

Prehistoric human colonization in the Pacific happened in two phases. Initially, Pleistocene hunter–gatherer expansions from Island Southeast Asia through New Guinea reached the Bismarck archipelago by 33,000 BP and the Papuan-speaking descendants of these people are dispersed throughout New Guinea and parts of Island Melanesia¹³. The second colonization wave of Austronesian language speakers involved a diaspora of Neolithic farming peoples out of south China and Taiwan around 6,000 BP^{13–15}. According to the ‘express train to Polynesia’ model, the Austronesian expansion from Taiwan was extremely rapid, taking roughly 2,100 years to reach the edges of western Polynesia—a distance of 10,000 kilometres.

Converging evidence from archaeology and molecular anthropology supports a rapid and relatively encapsulated dispersal of the Austronesian speakers throughout the Pacific^{13,16–18} (Fig. 1); however, there is some dispute about the exact degree of interaction with earlier Melanesian settlers, the rate at which the migration occurred and the extent and location of any colonization pauses¹⁹. In broad terms, most Pacific scholars seem to favour the express-train model, but others have argued that the ancestral Polynesians derive from an older Melanesian “matrix”^{7,20}. The latter authors stress that a phylogenetic, colonization-focused perspective obscures the high degree of prehistoric contact and inter-relationships amongst Pacific people; we use Terrell’s phrase⁶—the entangled-bank model—to represent this. These two models are not mutually exclusive, but are best characterized as two ends of a continuum of modes of human prehistory, with a pure tree at one end and a maximally connected network at the other. The issues surrounding the settlement of the Pacific are thus a microcosm of the general debate about whether human cultural evolution can be appropriately represented as a tree.

We tested one aspect of the express-train model, the colonization sequence, in the way that biologists test hypotheses about the sequence of events in biological evolution. We constructed a tree and then mapped the trait onto the tree to see whether the inferred sequence of changes fits a particular scheme²¹. Figure 2 shows how a simple colonization scheme can be tested by mapping geography onto an independent tree. We grouped languages according to Diamond’s archaeological/geographical stations²². Using character-state functions in the program MacClade²³, we assigned each station a character state from 0 to 9. The states were ordered in a character-state tree to fit the sequence proposed by the express-train model. For example, in Fig. 1 the Taiwanese languages were grouped as state 1, the Remote Oceanic languages as state 8; this means a change from state 1 to 8 would require five steps (according to the model presented in Fig. 1). By mapping these character states onto the most-parsimonious language tree (Fig. 3), we were able to evaluate the express-train model in a quantitative manner. If the language tree fits the express-train model well, then the character-state tree should fit well onto our obtained tree. The shortest possible tree length required to optimize the character-state tree onto the

language tree was nine (that is, the number of character states minus one). When the character-state tree was mapped onto the optimal tree, we obtained a tree length of 13. To assess the statistical significance of the fit, we randomly shuffled the character states between the 77 languages 200 times²³. This gave us a null distribution of tree lengths with a mean tree length of 48.9 steps (s.d. 1.98, range 43–53). This indicates that the express-train character-state tree fits the language tree with significantly fewer steps than would occur by chance. In fact, the obtained fit was very close to the shortest possible length (nine), indicating that the express-train model fits the language tree exceptionally well.

By definition, an entangled-bank model cannot be represented by a character-state tree; however, we can assess whether the language data support the entangled-bank model by examining the topology

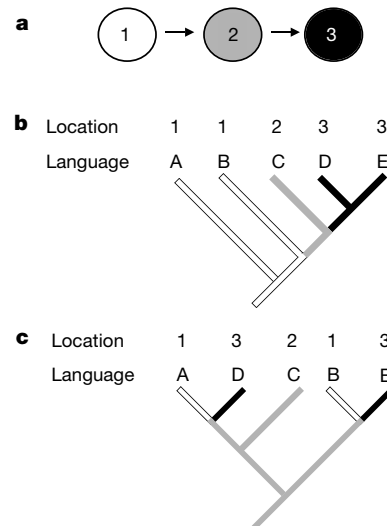


Figure 2 A phylogenetic approach to testing a colonization sequence. **a**, Model for the colonization of three areas, in which an ancestral population moves from area one to area two and then to area three. **b**, Tree that fits perfectly with the colonization model in **a** (fit = 2 steps). **c**, A tree that fits poorly with the colonization sequence (fit = 4 steps).

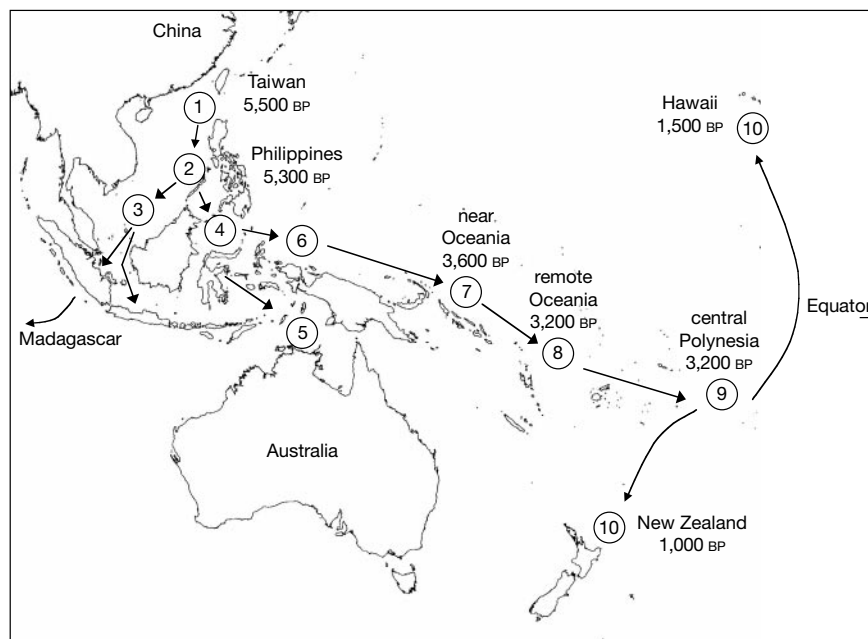


Figure 1 The express train to Polynesia model of the Austronesian colonization of the Pacific (adapted from refs 5 and 22). Approximate archaeological dates of settlement are indicated^{13,22}. Each ‘station’ is a separate character state: 1, Taiwan; 2, Philippines;

Chamorro, Palau; 3, Borneo, Indonesia, Malay; 4, Sulawesi; 5, central Malayo-Polynesian; 6, south Halmahera/west New Guinea; 7, near Oceania; 8, remote Oceania; 9, central Polynesia; 10, east Polynesia.

of the tree. While advocates of this model make no predictions about the likely shape of a language tree under an entangled-bank conception, they argue that large-scale migration patterns in languages are obscured by culture contact⁷. Consequently, they might predict a layered, 'candelabra-like' tree that emphasizes regional contact. In contrast, an (archaeologically) quick colonizing wave from Island Southeast Asia through the Pacific to Polynesia should produce a tree topology that is 'chain-like' (see Fig. 3). Proponents of the entangled-bank model argue that culture, language and biology 'combine and recombine' in such complex interactions that patterns of language relationships may tell us very little about the history of language speakers⁷. In this case, the tree should merely reflect geographical proximity. Our tree, however, shows several cases where the relationships fit the historical sequences implied by the express-train model but conflict with geographical proximity (see Fig. 3).

Although we reject the specific features of the entangled-bank

model, we do not claim that Austronesian cultural history is totally tree-like. The consistency index (a measure of the fit of the lexical data on the tree) is only 0.25. This value is not substantially lower than would be expected for equivalently sized morphological and molecular data sets²⁴ in which hybridization is uncommon. Although it is probable that much of the poor fit in the lexical data is due to the loss of cultural or linguistic features^{15,25}, archaeological²⁶ and genetic²⁷ evidence do indicate that population interaction and 'borrowing' are likely to have occurred even between far-flung archipelagoes. A way of approaching the issue of borrowing is to examine languages whose placement conflicts with the colonization scheme. For example, Buli and Numfor are grouped inside the Oceanic language group on our tree, whereas the express-train model places these south Halmahera/west New Guinea languages outside the Oceanic group. Similarly, Chamorro and Palau—languages whose closest relationships are most likely with the Philippines²⁸—are grouped with the Oceanic languages. In both these cases, borrowing is a likely cause of the incongruence between the express-train model and our tree. More detailed evidence for specific patterns of reticulation is evaluated elsewhere (F.M.J. and R.D.G. manuscript in preparation).

The patterns apparent in linguistic relationships are integrally tied to the movements, contacts and activities of language speakers. Our preliminary investigations have shown that a phylogenetic approach to languages offers the ability to test hypotheses about human prehistory. In biology, phylogenetic methods have become invaluable tools for investigating patterns and processes in evolution. In the future, phylogenetic methods may provide a common methodology and analytic framework to integrate data from ethnography, archaeology, linguistics and genetics. This is an important step towards a unified approach to biological and cultural evolution. □

Methods

Data were taken from Blust's Austronesian Comparative Dictionary (R. Blust, personal communication). This is a continuing project to compile comparative lexical data from the largest language family in the world. Currently, the dictionary is 25% complete and comprises 5,185 lexical items across 191 languages. Each lexical item has a set of cognate terms listed with the languages in which they appear. To ensure that there was sufficient information in the data set for phylogenetic analysis, we cut the number of languages from 191 to 68 by using a criterion of 150 or more appearances in a cognate set. An additional nine languages were then added to provide a balanced representation of the principal Austronesian language subgroups, giving us 77 languages in total. The presence of a language in a cognate set was coded as '1' in a matrix of 77 languages $\times$ 5,185 lexical items. If a language was not present in a particular cognate set, that language was coded as '0' for that item in the matrix. Linguistic^{15,28}, archaeological¹³ and genetic^{16,18} evidence agrees that Taiwan is the most likely Austronesian homeland, and so the two Taiwanese languages (Amis and Paiwan) were used to root the tree. We used PAUP* 4.0d65 (ref. 29) to find the set of most-parsimonious trees. To maximize the chance of finding optimal trees, 1,200 random addition sequences and tree bisection–reconnection branch swapping were used. Characters were typed as easy loss (5:1 ratio) on the assumption that independent losses of lexical items were more likely than independent gains. Similar assumptions about character coding have been used for complex behavioural characters³⁰, and linguistic features (such as phonemes) have been shown to be lost in a west-to-east direction across the Pacific²⁵. Other easy loss codings and equally weighted parsimony produced similar results (R.D.G. and F.M.J., manuscript in preparation). The search found one shortest tree of 52,129 steps with a consistency index of 0.25. The linguistic data set contained significant phylogenetic signal (treelength skewness index $g_1 = -0.505$ calculated from 100,000 random trees).

Received 6 January; accepted 28 April 2000.

1. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Oxford Univ. Press, Oxford, 1871).
2. Kirch, P. V. & Green, R. C. History, phylogeny, and evolution in Polynesia. *Curr. Anthropol.* **28**, 431–456 (1987).
3. Mace, R. & Pagel, M. The comparative method in anthropology. *Curr. Anthropol.* **35**, 549–564 (1994).
4. Ruvolo, M. in *Biological Metaphor and Cladistic Classification: An Interdisciplinary Perspective* (eds Hoenigswald, H. M. & Wiener, L. F.) 193–216 (Univ. Pennsylvania Press, Pennsylvania, 1987).
5. Diamond, J. M. Express train to Polynesia. *Nature* **336**, 307–308 (1988).
6. Terrell, J. History as a family tree, history as an entangled bank: Constructing images and interpretations of prehistory in the South Pacific. *Antiquity* **62**, 642–657 (1988).
7. Terrell, J., Hunt, T. L. & Gosden, C. The dimensions of social life in the Pacific. *Curr. Anthropol.* **38**, 155–195 (1997).
8. Hillis, D. M., Huelsenbeck, J. P. & Cunningham, C. W. Application and accuracy of molecular phylogenies. *Science* **264**, 671–676 (1994).
9. Crowley, T. *An Introduction to Historical Linguistics* 3rd edn (Oxford Univ. Press, Auckland, 1997).

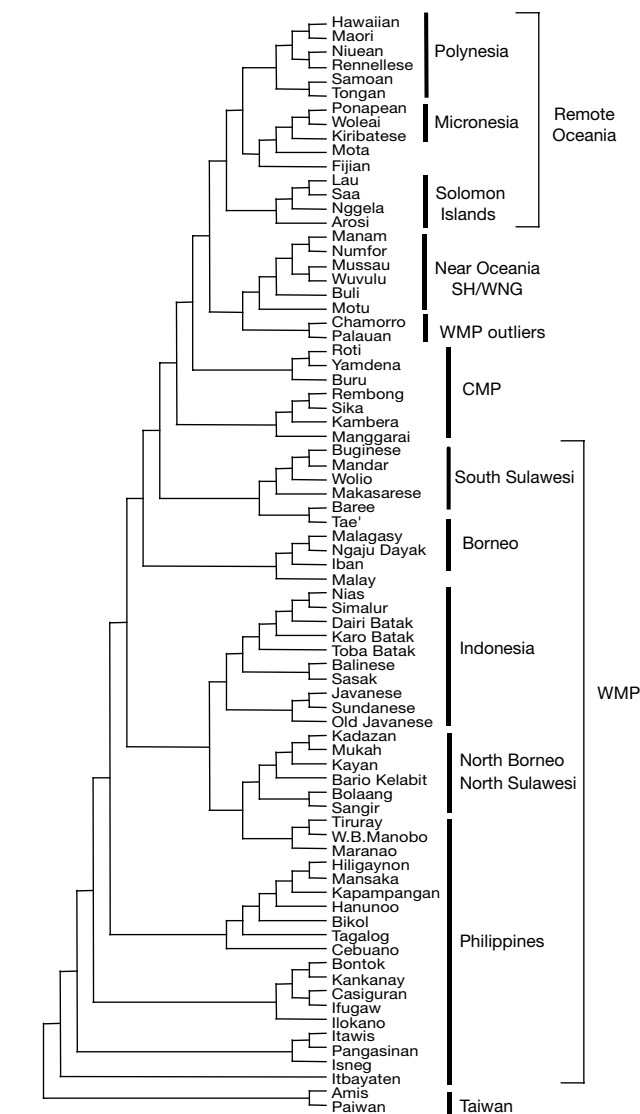


Figure 3 Phylogenetic tree of 77 Austronesian languages. WMP, Western Malayo-Polynesian; CMP, Central Malayo-Polynesian; SH/WNG, South Halmahera/West New Guinea. The topology of the tree shows considerable agreement with traditional linguistic groupings^{14,15,28}; these groupings reflect historical relationships, not just geographical proximity. For instance, Malagasy (spoken on Madagascar) is grouped with Ngaju Dayak from western Borneo. Tae' (from Central Sulawesi) groups within the south Sulawesi languages, whereas the north Sulawesi languages (for example, Bolaang Mongondow (Bolaang)) are more closely related to languages of north Borneo than are other Sulawesi languages.

10. Warnow, T. Mathematical approaches to comparative linguistics. *Proc. Natl Acad. Sci. USA* **94**, 6585–6590 (1997).
11. Bateman, R. et al. Speaking of forked tongues. *Curr. Anthropol.* **31**, 1–24 (1990).
12. Moore, J. H. Putting anthropology back together again: The ethnogenetic critique of cladistic theory. *Am. Anthropol.* **96**, 925–948 (1994).
13. Bellwood, P. *The Prehistory of the Indo-Malaysian Archipelago* 2nd edn (Univ. Hawaii Press, Honolulu, 1997).
14. Bellwood, P. The Austronesian dispersal and the origin of languages. *Sci. Am.* **265**, 88–93 (1991).
15. Blust, R. The prehistory of the Austronesian speaking peoples: a view from language. *J. World Prehist.* **9**, 453–510 (1995).
16. Melton, T. et al. Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am. J. Hum. Genet.* **57**, 403–414 (1995).
17. Green, R. C. Integrating historical linguistics with archaeology: Insights from research in Remote Oceania. *Bull. Indo-Pacific Prehist. Ass.* **18**, 3–16 (1999).
18. Lum, J. K. & Cann, R. L. mtDNA and language support a common origin of Micronesians and Polynesians in Island Southeast Asia. *Am. J. Phys. Anthropol.* **105**, 109–119 (1998).
19. Kirch, P. V. *The Lapita Peoples* (Blackwell, Cambridge, 1997).
20. Terrell, J. *Prehistory in the Pacific Islands: A Study of Variation in Language, Customs, and Human Biology* (Cambridge Univ. Press, Cambridge, 1986).
21. Martin, A. Hammerhead shark origins. *Nature* **364**, 494 (1993).
22. Diamond, J. M. *Guns, Germs and Steel* (Jonathan Cape, London, 1997).
23. Maddison, W. P. & Maddison, D. R. *MacClade: Analysis of Phylogeny and Character Variation Version 3.05*. (Sinauer Associates, Massachusetts, 1992).
24. Sanderson, M. J. & Donoghue, M. J. Patterns of variation in levels of homoplasy. *Evolution* **43**, 1781–1795 (1989).
25. Blust, R. in *Currents in Pacific Linguistics: Papers on Austronesian Languages and Ethnolinguistics in Honour of George W. Grace* (ed. Blust, R.) 27–42 (Australian National Univ., Canberra, 1991).
26. Weisler, M. I. Hard evidence for prehistoric interaction in Polynesia. *Curr. Anthropol.* **39**, 521–531 (1998).
27. Matisoo-Smith, E. et al. Patterns of prehistoric human mobility in Polynesia indicated by mtDNA from the Pacific rat. *Proc. Natl Acad. Sci. USA* **95**, 15145–15150 (1998).
28. Ross, M. D. in *Comparative Austronesian Dictionary: An Introduction to Austronesian Studies* (ed. Tryon, D. T.) 27–42 (ANU, Canberra, 1994).
29. Swofford, D. L. *Phylogenetic Analysis Using Parsimony (PAUP*) Version 4.0d65*. (Sinauer Associates, Massachusetts, 1999).
30. Cunningham, C. W., Omland, K. E. & Oakley, T. H. Reconstructing ancestral character states: a critical reappraisal. *Trends Ecol. Evol.* **13**, 361–366 (1998).

Acknowledgements

We are grateful to R. Blust for making the Austronesian Comparative Dictionary available to us. We thank M. Corballis, R. Green and A. Rodrigo for comments on the manuscript; R. Clark for advice; and H. Tse for programming assistance.

Correspondence and requests for materials should be addressed to R. D. G. (e-mail: rd.gray@auckland.ac.nz).

Mosaic evolution of brain structure in mammals

Robert A. Barton* & Paul H. Harvey†

* Evolutionary Anthropology Research Group, Department of Anthropology, University of Durham, Durham DH1 3HN, UK

† Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

The mammalian brain comprises a number of functionally distinct systems. It might therefore be expected that natural selection on particular behavioural capacities would have caused size changes selectively, in the systems mediating those capacities^{1–3}. It has been claimed, however, that developmental constraints limited such mosaic evolution, causing co-ordinated size change among individual brain components³. Here we analyse comparative data to demonstrate that mosaic change has been an important factor in brain structure evolution. First, the neocortex shows about a fivefold difference in volume between primates and insectivores even after accounting for its scaling relationship with the rest of the brain. Second, brain structures with major anatomical and functional links evolved together independently of evolutionary change in other structures. This is true at the level of both basic brain subdivisions and more fine-grained functional

systems. Hence, brain evolution in these groups involved complex relationships among individual brain components.

Studies of mammalian brain evolution have highlighted the neocortex as a structure associated with intelligence and flexible behaviour, which varies enormously in size between species^{4–6}. Large-brained mammals, such as primates, tend to have a neocortex that is disproportionately expanded relative to other structures³. The extent to which this size variation can be explained by allometric scaling relative to the rest of the brain, as opposed to size changes independent of other brain structures, remains unclear however^{3,7}. Figure 1 indicates clearly that neocortex size varies even after accounting for its scaling relationship with the size of the rest of the brain. The three parallel lines with different intercepts indicate taxonomic differences (grade shifts) in relative neocortex size between primates and insectivores, and, within the primates, between strepsirhine and haplorhine sub-orders. Independent contrasts analysis confirms the presence of significant grade shifts in relative neocortex size. First, the slopes are statistically indistinguishable (haplorhine versus strepsirhine primates: $t = 1.6$, degrees of freedom, d.f. = 37, $P = 0.13$; primates versus insectivores: $t = 0.6$, d.f. = 71, $P = 0.54$). Second, the absolute values of the contrasts between orders and sub-orders are unusually large and beyond the range of all other contrasts in each data set (haplorhine versus strepsirhine residual = 2.8 standard deviations greater than the mean; primate versus insectivore residual = 5.6 standard deviations greater than the mean). On the basis of separate regression equations for insectivores and primates (averaging between strepsirhines and haplorhines), a primate with a non-neocortical brain size of 1,000 mm³ would have a neocortex nearly five times larger than would an insectivore with the same non-neocortical brain size (881 mm³ versus 187 mm³). In some specific cases, we observe an even greater difference in relative size. For example, the common tenrec *Tenrec ecaudatus*, an insectivore, has a non-neocortical brain volume somewhat greater than that of the

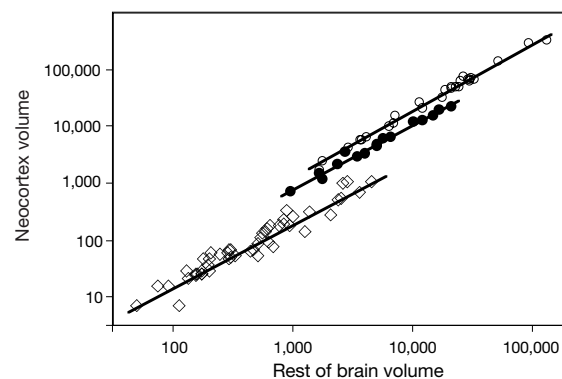


Figure 1 Taxonomic differences in relative neocortex size among primates (strepsirhines and haplorhines) and insectivores. Brain part volumes are in cubic millimetres. Open circles, haplorhine primates; closed circles, strepsirhine primates; diamonds, insectivores. Slopes (and 95% confidence intervals) for insectivores, strepsirhines and haplorhines respectively are 1.11 (1.03–1.20), 1.13 (1.04–1.22) and 1.20 (1.14–1.26).

Table 1 Regression statistics for the scaling of neocortical white and grey matter volume on volume of the rest of the brain

	White matter volume			Grey matter volume		
	Slope	Confidence intervals	r ²	Slope	Confidence intervals	r ²
Insectivores	1.32	1.23–1.41	0.95	1.09	0.94–1.18	0.94
Strepsirhines	1.48	1.32–1.65	0.99	1.06	0.98–1.14	0.99
Haplorhines	1.53	1.37–1.67	0.98	1.12	1.07–1.18	0.99
New World Monkeys	1.40	1.20–1.59	0.98	1.08	0.96–1.21	0.98
Old World monkeys	1.42	0.13–2.71	0.92	0.97	0.45–1.49	0.97

marmoset *Cebuella pygmaea* (2,058 versus 1,770 mm³), yet the marmoset's neocortex is nearly ten times larger (2,535 versus 273 mm³). Substantial taxonomic differences in neocortex size thus exist after taking scaling into account. Although the rest of the brain includes the olfactory bulb, which has reduced in primates, the taxonomic differences in neocortex size are still apparent when scaled against non-olfactory structures.

How important are these grade shifts, as opposed to allometric scaling, in explaining the disproportionate expansion of the neocortex in large-brained taxa? Figure 1 indicates that, when the effects of the grade shifts between orders and sub-orders are taken into account, the scaling of neocortex size is nearly isometric (that is, in direct proportion to the rest of the brain, as would be indicated by a slope between log-transformed variables of 1). This indicates that allometric scaling has a relatively small effect on proportional differences in neocortex size. Furthermore, the slight departure from isometry is attributable to the white matter component alone (Table 1). White matter volume of the neocortex scales with marked hyper-allometry relative to the rest of the brain. The white matter consists of fibres connecting neocortical areas to each other and to other structures, and the hyper-allometry is predicted by simple

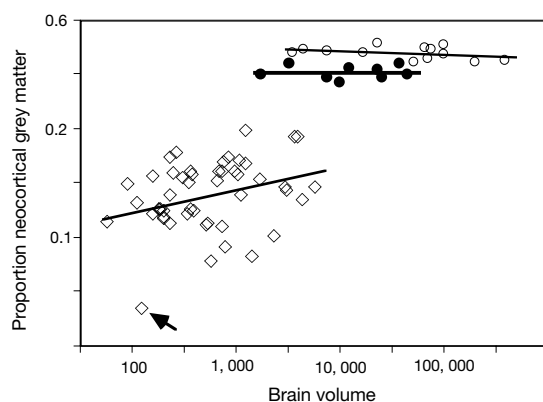


Figure 2 Proportion of brain volume composed of neocortical grey matter in relation to overall brain volume. Symbols as in Fig. 1. The proportion of grey matter is uncorrelated with brain volume in strepsirrhines ($r^2 = 0.0004$, $n = 9$, $P = 0.99$), haplorhines ($r^2 = 0.07$, $n = 13$, $P = 0.20$) or the combined primate sample ($r^2 = 0.01$, $n = 22$, $P = 0.28$). The weak positive correlation for insectivores ($r^2 = 0.08$, $n = 48$, $P = 0.03$) becomes non-significant when the outlying species (*Geogale aurita*, arrowed) is removed ($r^2 = 0.04$, $n = 47$, $P = 0.09$). Independent contrasts analysis confirms the lack of association between brain size and proportion of neocortical grey matter (primates: $r^2 = 0.01$, $n = 20$, $P = 0.72$; insectivores: $r^2 = 0.05$, $n = 32$, $P = 0.12$).

Table 2 Correlated volumetric evolution among major brain structures revealed by multiple regressions on independent contrasts

Primates ($n = 40$ independent contrasts)				
	Diencephalon	Mesencephalon	Cerebellum	Medulla
Neocortex	0.65 (4.06***)	0.05 (0.31)	0.51 (4.54****)	-0.21 (1.80)
Diencephalon		0.43 (3.42**)	0.03 (0.21)	0.13 (1.22)
Mesencephalon			0.02 (0.11)	0.35 (3.15**)
Cerebellum				0.32 (2.33*)
Insectivores ($n = 33$ independent contrasts)				
	Diencephalon	Mesencephalon	Cerebellum	Medulla
Neocortex	1.12 (4.94****)	-0.19 (0.96)	0.21 (1.24)	-0.16 (0.99)
Diencephalon		0.38 (3.83***)	0.23 (2.35*)	-0.02 (0.88)
Mesencephalon			-0.26 (1.70)	0.52 (4.44****)
Cerebellum				0.43 (2.72**)

Standardized regression coefficients are given with associated t -values and significance levels in parentheses. Structures in the left column were regressed on those in the top row (results are given only above the diagonal because identical t values and P values are obtained when regressing structures along the top row on structures in the left column). Significant t values indicate that the two structures exhibit significantly correlated evolution with the effects of change in the other structures partialled out. ****, $P < 0.0001$; ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$. The patterns of correlated evolution are displayed visually in Fig. 3.

scaling models of connectivity in brains of varying size^{8,9}. In contrast, grey matter volume scales isometrically—a conclusion that holds when neocortical grey matter volume is scaled against non-cortical nuclei which themselves have no major white matter component (unpublished analysis of cerebellar and medullary nuclei). Hence, taxonomic differences in the proportion of the brain composed of neocortical grey matter volume are independent of overall brain size (Fig. 2).

Next we examine the patterns of interrelationships among brain structures. If brain evolution occurred by size change concentrated in specific brain systems, significant correlations should be found between structures linked by important functional and anatomical connections, even after accounting for the effects of size change in other structures. We analysed correlated evolution among brain structures in primates and insectivores, using the method of independent contrasts. The first analysis was restricted to major structures of the mammalian brain, the neocortex, diencephalon, mesencephalon, cerebellum and medulla, to assess whether mosaic evolution is detectable even at this anatomically crude level. The neocortex was included rather than the entire telencephalon, as the latter incorporates structures that vary greatly in functional and connectational properties (some of these are tested in the subsequent analyses). The results (Table 2) are summarized in Fig. 3, and give rise to two main conclusions. First significant partial correlations exist. That is, particular pairs of structures show significantly correlated volumetric evolution even after accounting for the effects of change in the other structures. Second, the patterns of correlated evolution are strikingly similar in the two orders. Of the five significant positive partial correlations in each taxon, four are shared between taxa. The explanation for this resemblance could be fundamental similarities in anatomical and functional connections, common developmental constraints^{3,7}, or both.

Whichever explanation is correct, the patterns of covariation indicate that, even at this anatomically crude level, brain structure evolution involved a complex set of relationships among individual structures. The chain of structures along the main axis in Fig. 3 (medulla–mesencephalon–diencephalon–neocortex) corresponds to a basic anatomical sequence from posterior (medulla) to anterior (neocortex) parts of the brain, and major projections are found between each of the links in this chain¹⁰. For example, the main part of the diencephalon, the thalamus, is the site of many major relays to and from the neocortex. The only difference between the patterns for the two orders is that cerebellum size correlates with neocortex size in primates, and with diencephalon size in insectivores. Extensive connections exist between neocortex and cerebellum. Given that many of these are relayed through the diencephalic thalamus,

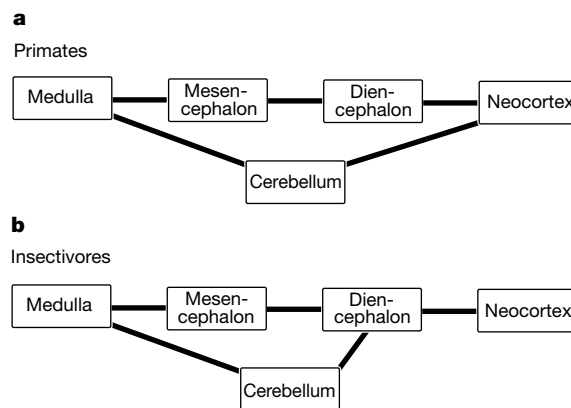


Figure 3 Correlated evolution among major brain structures. **a**, Primates. **b**, Insectivores. Connecting lines indicate significant positive partial correlations between the structures, from the analyses in Table 2. Hence, structures connected by such lines have evolved together independently of evolutionary change in the other structures.

there is likely to be a tight three-way evolutionary relationship amongst neocortex, cerebellum and diencephalon. This idea is supported by repeating the two analyses for the cerebellum with its primary correlate (either neocortex size or diencephalon size) removed. As predicted, cerebellum size in insectivores correlates significantly with neocortex size when diencephalon size is excluded ($P < 0.0001$), and in primates cerebellum size correlates significantly with diencephalon size when neocortex size is excluded ($P = 0.003$).

The major brain sub-divisions in Fig. 3 are anatomically and functionally heterogeneous, and analysis of smaller, more functionally homogenous sub-divisions might reveal more fine-grained patterns of correlated evolution. We therefore tested for correlated evolutionary change in the individual components of specific functional units or systems. Data are available for six different systems, five in each order, yielding ten separate tests in all. The results (Fig. 4) reveal that, in all ten cases, components of functional systems evolved together independently of evolutionary size change in the other structures and in the rest of the brain. Hence, components of functional systems have highly specific evolutionary relationships, not attributable to their membership of more global systems such as the whole brain or the limbic system. In addition, although six of the tests also reveal a significant correlation with one of the other structures, in nine out of the ten cases the predicted relationship is the strongest (the only exception being where a structure—the insectivore hippocampus—correlates more

strongly with the rest of the brain, perhaps reflecting the many diverse outputs and inputs of the hippocampus). Furthermore, some additional relationships are of course predicted by the hypothesis that functionally linked structures evolved together. Only two significant relationships out of the 48 tested seem anomalous in neurobiological terms, and these two can be reduced to a single case, as they involve the same two systems (amygdala and vestibular–cerebellar system in primates). The other significant results are explicable in functional terms. In both primates and insectivores there is a correlation between amygdala and olfactory components, reflecting the close anatomical and functional links between these structures^{10–12}. In primates, the negative correlation between the lateral geniculate nucleus, a visual structure, and the olfactory bulb, may indicate a trade-off between visual and olfactory sensory modalities, perhaps associated with divergence into nocturnal versus diurnal niches¹³. The most striking feature of Fig. 4, however, is the combination of highly significant relationships within functional systems and the general absence of such relationships between systems.

These comparative analyses show that mammalian brain evolution involved size changes concentrated in specific structures and functional systems. One implication is that, as in birds¹⁴, the cognitive and ecological significance of species differences in brain size should be evaluated by examining which neural systems in particular have been the target of selection. Although there may be some constraints on evolutionary change in individual neural

a Primates						
	Rest of brain	1. Olfactory bulb	2. Entorhinal cortex	3. CBL of amygdala	4. Cerebellar nuclei	5. Visual cortex
1. Olfactory cortex	0.18 0.74	0.31 3.43**	0.09 0.67	0.32 2.13*	0.23 1.11	---
2. Hippocampus	0.37 1.46	-0.06 -0.62	0.61 4.23***	-0.12 -0.77	0.14 0.66	---
3. CM of amygdala	-0.36 -1.15	0.12 1.04	-0.30 -1.71	0.91 4.96****	0.65 2.57*	---
4. Vestibular nuclei	-0.38 -1.40	0.13 1.28	-0.11 -0.69	0.35 2.16*	0.95 4.27****	---
5. Lateral geniculate	1.04 1.49	-0.52 -2.1*	0.49 1.27	-0.40 -1.05	-0.70 -1.33	0.83 2.33*

b Insectivores						
	Rest of brain	1. Olfactory bulb	2. Entorhinal cortex	3. CBL of amygdala	4. Cerebellar nuclei	5. Superior olive
1. Olfactory cortex	-0.05 -0.59	0.85 10.72****	0.18 1.87	0.05 0.42	---	---
2. Hippocampus	0.68 4.70****	0.12 0.90	0.34 2.22*	-0.12 -0.61	---	---
3. CM amygdala	0.16 1.30	0.28 2.46*	0.01 0.10	0.54 3.19**	---	---
4. Vestibular nuclei	1.41 0.18	-0.07 -0.33	0.25 0.85	0.04 0.14	0.65 2.36*	---
5. Cochlear nuclei	0.11 0.48	0.07 0.51	-0.27 -1.36	-0.05 0.24	---	1.03 10.37****

Figure 4 Correlated volumetric evolution of functionally related brain structures. Each row summarizes results of one multiple regression based on independent contrasts. Contrasts in the volume of each structure in the left column were regressed on volumes of structures in the top row. In each cell: top figure, standardized regression coefficient; bottom figure, *t*-value; asterisks, level of significance, as in Table 2. A significant result means that the two structures evolved together independently of evolutionary change in the other

structures for which results are reported in the same row. Predicted relationships are indicated by the bold boxes. These are the relationships between pairs of structures that are components within a functional system. The systems are 1, olfactory system; 2, hippocampal formation; 3, amygdala; 4, sensory-motor (vestibular) system; 5, visual system (primates) or auditory system (insectivores). CM, centro medial complex; CBL, cortico-basolateral complex; entorhinal cortex includes subiculum.

systems, tending to result in coordinated evolution among the majority of brain structures^{3,7}, such constraints are evidently insufficiently tight to prevent the type of system-specific change documented here. We conclude that mosaic evolution has been an important factor in the adaptive radiation of the mammalian brain. □

Methods

All volumetric measurements were made by the same research group using uniform methods^{15–19}. Scaling relationships among biological variables are generally well described by the power function:

$$Y = kX^\alpha \quad (1)$$

$$\log Y = \alpha(\log X) + \log k \quad (2)$$

where Y and X are the variables, α and k are the parameters of the power equation. With logarithmic transformation the relationship becomes linear, so that the exponent is expressed as the slope of the line:

Slopes were determined using least-squares regression. As coefficients of determination in all scaling analyses were uniformly high (>0.92), use of alternative line-fitting techniques would have minimal impact. To show the presence of taxonomic grade shifts clearly, we based graphs on values for individual species, rather than on independent contrasts (see below). A grade shift is defined as a taxonomic difference in the mean value of a continuous variable after the effects of scaling have been taken into account.

To verify statistically the existence of significant grade shifts among taxa, we used a procedure based on the method of independent contrasts^{20–22}. This procedure involves (1) demonstration that the slopes for the different taxa are homogenous, based on a t -test on the residuals from a regression of independent contrasts for the dependent variable on contrasts for the scaling variable; (2) demonstration that the contrast between the taxa being compared is an outlier, and hence has an unusually large residual value compared with other contrasts in the data set²². The program and procedures for generating independent contrasts have been described²², and the phylogenies of primates and insectivores are from ref. 23 and R. Grenyer (personal communication), respectively.

Independent contrasts were also used to test for correlated evolution among brain structures. Assumptions of the method were checked as described in ref. 22. In these analyses, data for strepsirrhine and haplorhine primates were pooled to yield adequate sample size. In the analyses presented in Table 2 (summarized in Fig. 3), independent contrasts in the volume of each structure were tested against the other structures in separate multiple regressions. This is equivalent to testing for significant partial correlations amongst the five structures, with the advantage that the regressions can be set through the origin as required with independent contrasts. A similar procedure was used in the analyses of correlated evolution among sub-components of functional systems (Fig. 4). We made predictions about which specific structures were likely to have evolved together, on the basis of well-known neurobiological links, and tested for correlated evolution between them, controlling for variation in the rest of the brain. We tested such predictions for pairs of structures within six systems, the olfactory system (olfactory cortex and olfactory bulb), visual system (primary visual cortex and lateral geniculate nucleus), auditory system (cochlear nuclei and superior olive), a sensory–motor (vestibular) system (vestibular nuclei of the medulla and cerebellar nuclei), the hippocampal formation (hippocampus and entorhinal cortex plus subiculum) and the amygdala (cortico-basolateral and centro-medial amygdala). These were chosen because of the close and relatively uncontroversial anatomical and functional links between the structures within each system, and because of the availability of comparative volumetric data. To assess how specific the relationships among components of each functional system were, we included components of the other systems as independent variables in the multiple regressions. However, because relatively few data were available for the primate visual system, the insectivore auditory system and insectivore vestibular system, components of these were excluded as independent variables in the analyses of other systems. In each case, variation in the volume of the other structures was taken into account using multiple regression. Olfactory cortex volume in primates was estimated as palaeocortex minus amygdala, thus including the lateral olfactory tract, olfactory tubercle and fibres of passage^{15,16}.

Received 1 February; accepted 4 May 2000.

- Eisenberg, J. *The Mammalian Radiations* (University of Chicago, Chicago, 1981).
- Harvey, P. H. & Krebs, J. R. Comparing brains. *Science* **249**, 140–146 (1990).
- Finlay, B. L. & Darlington, R. B. Linked regularities in the development and evolution of mammalian brains. *Science* **268**, 1578–1584 (1995).
- Innocenti, G. M. & Kaas, J. H. The cortex. *Trends Neurosci.* **18**, 371–372 (1995).
- Kaas, J. H. The evolution of isocortex. *Brain Behav. Evol.* **46**, 187–196 (1995).
- Dunbar, R. I. M. The social brain hypothesis. *Evol. Anth.* **6**, 178–190 (1998).
- Finlay, B. L., Darlington, R. B. & Nicastro, N. Developmental structure in brain evolution. *Behav. Brain Sci.* (in the press).
- Hofman, M. A. On the evolution and geometry of the brain in mammals. *Prog. Neurobiol.* **32**, 137–158 (1989).
- Ringo, J. L. *Brain. Behav. Evol.* **38**, 1–6 (1991).
- Butler, A. B. & Hodos, W. in *Comparative Vertebrate Neuroanatomy: Evolution and Adaptation*. (Wiley-Liss, New York, 1996).
- Cousens, G. & Otto, T. Both pre- and post-training excitotoxic lesions of the basolateral amygdala abolish the expression of olfactory and contextual fear conditioning. *Behav. Neurosci.* **112**, 1092–1103 (1998).
- Swanson, L. W. & Petrovich, G. D. What is the amygdala? *Trends Neurosci.* **21**, 323–331 (1998).
- Barton, R. A., Purvis, A. & Harvey, P. H. Evolutionary radiation of visual and olfactory brain systems in primates, bats and insectivores. *Phil. Trans. Roy. Soc. B* **348**, 381–392 (1995).
- Krebs, J. R. Food-storing birds: adaptive specialization in brain and behaviour? *Phil. Trans. Roy. Soc. B* **329**, 153–60 (1990).
- Stephan, H., Frahm, H. D. & Baron, G. New and revised data on volumes of brain structures in insectivores and primates. *Folia Primatol.* **35**, 1–29 (1981).
- Stephan, H., Baron, G. & Frahm, H. D. in *Comparative Brain Research in Mammals. Vol. 1: Insectivores* (Springer, New York, 1991).
- Frahm, H. D., Stephan, H. & Stephan, M. Comparison of brain structure volumes in Insectivora and Primates. I. Neocortex. *J. Hirnforsch.* **23**, 375–389 (1982).
- Matano, S., Baron, G., Stephan, H. & Frahm, H. D. Volume comparisons in the cerebellar complex of primates. II Cerebellar nuclei. *Folia primatol.* **44**, 182–203 (1985).
- Matano, S. A volumetric comparison of the vestibular nuclei in primates. *Folia primatol.* **47**, 189–203 (1986).
- Felsenstein, J. Phylogenies and the comparative method. *Am. Nat.* **125**, 1–15 (1985).
- Harvey, P. H. & Pagel, M. D. *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, Oxford, 1991).
- Purvis, A. & Rambaut, A. Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data. *Comp. Appl. Biosci.* **11**, 247–251 (1995).
- Purvis, A. A composite estimate of primate phylogeny. *Phil. Trans. Roy. Soc. B* **348**, 405–421 (1995).

Acknowledgements

We thank R. Grenyer for permission to use an unpublished phylogeny of insectivores.

Correspondence and requests for materials should be addressed to R.A.B. (e-mail: r.a.barton@durham.ac.uk).

The temporal response of the brain after eating revealed by functional MRI

Yijun Liu*, Jia-Hong Gao†, Ho-Ling Liu† & Peter T. Fox†‡

* Department of Psychiatry and Brain Institute, University of Florida, 100 Newell Drive, Gainesville, Florida 32610, USA

† Research Imaging Center and ‡ Department of Physiology, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, Texas 78284, USA

After eating, the human brain senses a biochemical change and then signals satiation, but precisely when this occurs is unknown. Even for well-established physiological systems like glucose–insulin regulation, the timing of interaction between hormonal processes and neural events is inferred mostly from blood sampling^{1–6}. Recently, neuroimaging studies have provided *in vivo* information about the neuroanatomical correlates of the regulation of energy intake^{7–10}. Temporal orchestration of such systems, however, is crucial to the integration of neuronal and hormonal signals that control eating behaviour¹¹. The challenge of this functional magnetic resonance imaging study is to map not only where but also when the brain will respond after food ingestion. Here we use a temporal clustering analysis technique to demonstrate that eating-related neural activity peaks at two different times with distinct localization. Importantly, the differentiated responses are interacting with an internal signal, the plasma insulin. These results support the concept of temporal parcellation of brain activity¹², which reflects the different natures of stimuli and responses. Moreover, this study provides a neuroimaging basis for detecting dynamic processes without prior knowledge of their timing, such as the acute effects of medication and nutrition in the brain.

The timing for the regulation of food ingestion is different from that for the control of sensorimotor or cognitive tasks. The question addressed here is: using functional magnetic resonance imaging

(fMRI), how can we map the brain dynamically if we do not know when there will be a response? Unlike electrophysiological methods (for example, intracortical electrical recording or EEG, which have a limited number of signal sources), fMRI deals with tens of thousands of voxels, which makes it impractical to show the time courses of signals from all spatial loci. Although most fMRI studies rely on

an explicit temporal profile that is correlated with a presumed haemodynamic function, the actual form of the fMRI response is unknown^{13–15}. The situation becomes more complicated when we need to image processes like memory consolidation, mood induction and drug intervention, for which the exact timing of the brain response is not known. In the experiment reported here (Fig. 1), the subjects underwent a continuous functional scan for 48 min, ingesting glucose at 10 min. Thus, for a 38-min post-ingestion period, there was no time window or preset model available for using the model-dependent functional mapping techniques^{16,17}. Even knowing that the hypothalamus might be involved, without temporal information it was difficult to define the region of interest within the hypothalamus from which a time course could be extracted.

To address this, we developed temporal clustering analysis (TCA), a method for searching for the maximal response in a combination of signal intensity and spatial extent (Fig. 2). TCA is based on a probability distribution of the overall brain voxels that reach the maximal signal intensity change. Note that the three-dimensional (3D) brain was collapsed into a 1D space, that is, the number of voxels, $N_{\max}$. Mathematically, TCA converts a multiple-dimension data space (x, y, z, t, S) into a simple relationship between the number of voxels reaching maximum and the time, to give the temporal maxima. $N_{\max}$ is a function of time (Fig. 2a), so that at a certain moment, $N_{\max}$ voxels reach a maximal signal. We found two peaks of response after glucose intake. The first peak immediately followed the stimulation, and the second one, which may be related to the plasma glucose changes, was delayed about 10.3 min.

Figure 2b shows a control experiment without ingestion of anything. Here, the temporal maxima of the fMRI signal are randomly or more evenly distributed compared with Fig. 2a. This implies that, without stimulation or disturbance, the resting brain activity may stay at a baseline level embedded in both scanner and physiological noises. When a strong stimulus was introduced during the scan, the balance was disturbed and the changes in brain activity resulted in the clustering of the maximal responses across time. In other words, although many brain regions could be involved in the whole process of glucose intake, at certain moments some of these spatially distributed regions may be tuned to respond maximally to the physiological changes. These regions form temporal clusters of overall voxel numbers that are significantly above the noise level.

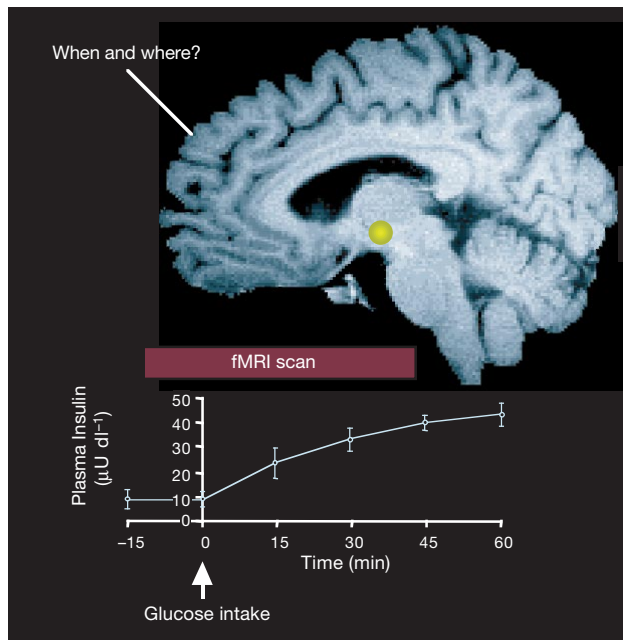


Figure 1 Neuro-biochemical interaction: an fMRI protocol for tracing the time course of brain activation following oral glucose intake. Using a high-resolution (1 mm) gradient-echo MRI sequence, a mid-sagittal section was scanned before, during and after glucose administration. Ten minutes of baseline (fasting-state) images were acquired before glucose intake, with continuous imaging for another 38 min after the stimulation. Blood was sampled before, during and after glucose ingestion, at 15-min intervals. The time course of the plasma insulin concentration (mean $\pm$ s.d., $n = 18$) is shown to demonstrate a potential interaction between the changes in the biochemical signals and in hypothalamic activity.

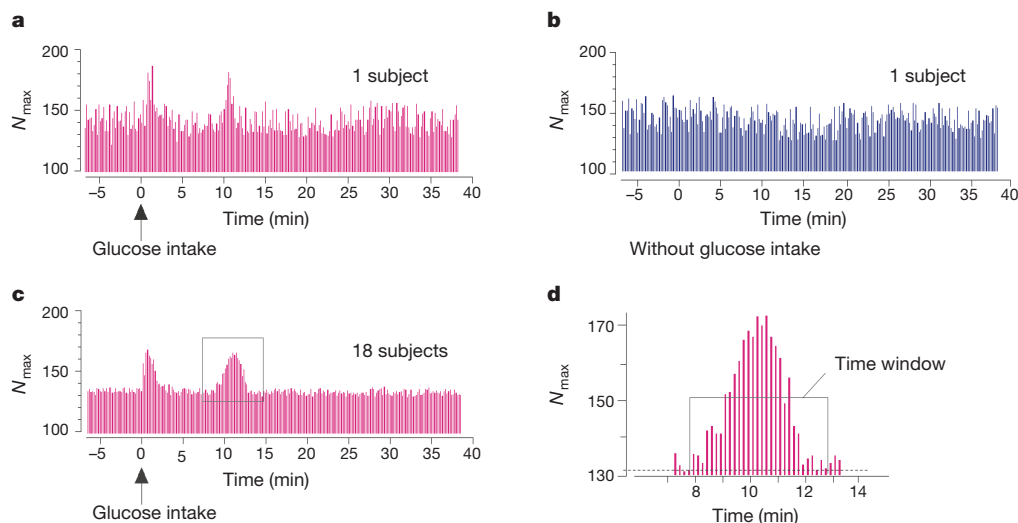


Figure 2 Temporal maxima of brain activity following glucose ingestion. Temporal clustering analysis (TCA) was used to track the maximal brain response, $N_{\max}$. **a**, Temporal maxima in 1 of 18 subjects with glucose intake; **b**, In 1 of 5 subjects without ingesting anything (other procedures remained same); and **c**, averaged over 18 subjects with

glucose intake. **d**, The zoomed distribution of temporal maxima around the second peak as framed in **c**. A time window T_w (7.7–12.8 min), centred on $t_{\max}$ (10.3 min), was determined by gaussian fitting. Dashed line, baseline level of the temporal maxima.

Although the distribution of the maximal signals was subject to random noise in a single subject (Fig. 2a, b), averaging of the temporal clusters pooled from eighteen subjects suppressed the variations (Fig. 2c, d). As TCA is based on voxel number, it is independent of the anatomical features in each individual brain. Therefore, this method may avoid the structural variability between individuals and the cross-subject average becomes more reliable, regardless of the actual activation loci and their variation among subjects. Two temporal clusters (or peaks) of the response are more distinct in the group data (Fig. 2c) than in the single subject data (Fig. 2a, b). Furthermore, the average results accommodated inter-individual variation in the timing of maximal responses, as seen in the distribution. However, the distribution itself was not a time course of the fMRI signal only drawn from the activated voxels. Rather, the activation of such voxels lay above the whole brain noise. Note that the time courses could be different among those activated voxels, and some of them might be immersed in the noise. The important thing here is that TCA first records the maximal responses for creating a time window to locate the functional regions, from which the actual time courses can then be drawn.

On the basis of the pooled temporal maxima, we identified a time window from the distribution around the peak (Fig. 2d). We used a

gaussian function to fit the data and extract the temporal profile of the brain activation. Two time windows were defined: 1.1–2.7 min for the first peak and 7.7–12.8 min for the second peak. Although the moment of maximal response could be measured for each subject (Fig. 2a), mapping the location still relied on the temporal profile, from which statistical tests isolated activated voxels from noise. To a first approximation, the spatial functional mapping was performed by comparing the mean of the voxel-wise fMRI signals from the defined time windows with that from the baseline.

Figure 3 shows the dynamic mapping of the brain responses. At the first peak, activated regions were found in the supplementary motor area (SMA), somatosensory cortex, cerebellum, anterior cingulate and orbitofrontal cortex (Fig. 3a, b). We initially speculated that activation of the SMA and cerebellum were responsible for sensorimotor activities related to the eating process itself. However, when we used pure water as a control stimulus, we found no activation in the SMA or cerebellum ($n = 8$), despite some slight activation in the somatosensory cortex and the upper-anterior region of hypothalamus during the early time frame. These results indicate that the early activation induced by glucose ingestion was not due to swallowing movements or other motion artefacts during eating. Rather, the SMA and other activated regions (Fig. 3a) may be involved in the integration of sensory and visceral signals, and affective activity associated with appetite and taste or olfaction^{8,18–20}.

At the second peak (Fig. 3c, d), we found a negative response in the medial hypothalamus, a brain structure that has been implicated in controlling feeding behaviour and regulating the plasma glucose concentration^{3–6}. There are two possible explanations for this negative response. First, there was enhanced baseline activity in the hypothalamus after a fasting period, and the glucose brought down this baseline. Second, the glucose activated inhibitory pathways to the hypothalamus, causing a decrease in hypothalamic activity. The exact neurophysiological mechanisms underlying the negative response, however, are unknown.

Further analyses revealed the most important results of this study: there is a dynamic interaction between the fMRI response and biochemical signal, specifically, the plasma insulin (Fig. 4). The data showed that the level of delayed activity in the hypothalamus is significantly correlated with the fasting plasma insulin concentrations (Fig. 4b). In contrast, no correlation was found between the early activation in the sensorimotor cortex and insulin concentration (Fig. 4a). The hypothalamus is important in the regulation of the plasma glucose concentration by modulating insulin secretion^{3,4}. Therefore, the correlation between the fMRI response and insulin amount may indicate a specific relationship between the regulation of glucose ingestion and the delayed activation in the medial hypothalamus. Combined with Fig. 3, these data indicate a dynamic dissociation of brain activity, in which both the timing and the spatial locations of the responses can be differentiated. The

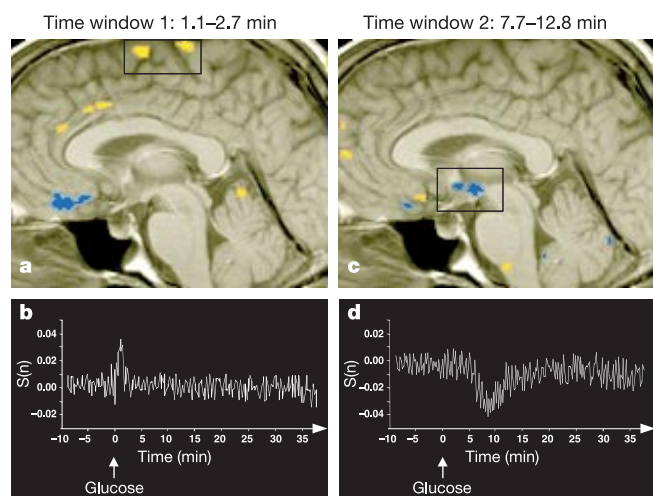


Figure 3 Mapping of dynamic brain activities following glucose ingestion. **a**, Activation during the first time window. **b**, Time course of the activation in one subject, drawn from an activated voxel in the ROI delineated in the sensorimotor cortex (**a**). **c**, Activation during the second window. **d**, Time course in the hypothalamic ROI in the same subject (**c**). A pixel-clustering size of 5 and a z -threshold of $|z| > 3.0$ were chosen to afford a $P < 0.01$ level of statistical significance of the detected signal changes^{12,17}. Yellow, significant increase; blue, significant decrease in the fMRI signal.

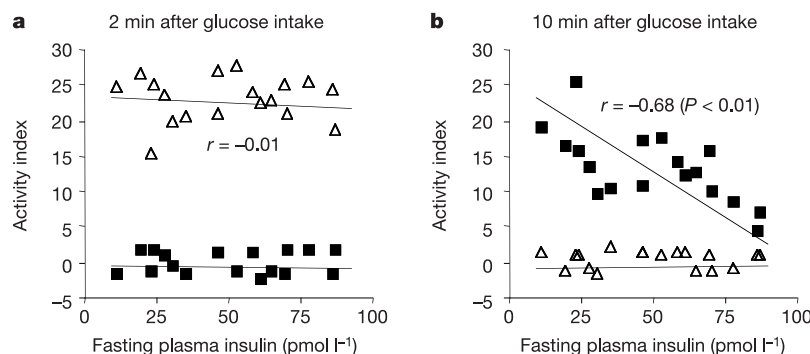


Figure 4 Time-dependent correlation between the fMRI response and the fasting plasma insulin concentrations. **a**, During the first time window. **b**, During the second time window. The cross-subject correlation (r) was calculated by a linear regression between the fMRI activity index¹², and the insulin concentration measured by blood sampling before glucose

intake. The significance of the correlation (r versus zero, $n = 18$) was determined using a one-tailed Student's t -test. Triangles, region of interest delineated in the sensorimotor cortex (Fig. 3a); squares, region of interest delineated in the hypothalamus (Fig. 3c).

localized activations are consistent with previous findings of the neural correlates underpinning the control of eating behaviour^{7–10}. Although the exact role in regulating food ingestion remains to be established in humans^{1,20}, the temporal information may help to understand how our brain generates a signal of satiety, both physiologically and psychologically^{7–10,21,22}.

As this fMRI study was limited to the imaging of a single brain slice, the current data should be interpreted with caution. The temporal maxima of the response could be different between single slice data and whole brain data, which may include high levels of effects in the prefrontal cortex^{8,21,22}. Many other physiological variables, such as changes in cardiac rate and blood pressure, may also influence the temporal maxima distribution. However, TCA has an advantage in that it is independent of brain anatomy, allowing the temporal profile to be extracted by pooling the data from a population of subjects. Therefore, by averaging, the major effects of glucose intake can be potentially distinguished from minor or random effects. In addition, TCA can be applied to whole brain fMRI studies²³ because the basic concept about the underlying neurobiology is the same; brain activity is both spatially and temporally parcelled¹².

Brain activation forms spatial clusters or local maxima^{24–27}, defined as the number of isolated regions with a peak intensity above a set threshold using the theory of gaussian random fields²⁸. On the basis of equivalent assumptions^{26–28}, this study has included the temporal extent (adding a new dimension) in the clustering analysis. The spatial maxima reflect the cellular parcellation or the Brodmann cytoarchitectonic map of the human brain. However, as shown here and in previous studies using sensorimotor and cognitive tasks^{12,29}, the partitioned brain functions implicate a temporal parcellation of neuronal activities. Therefore, a brain mapping method needs to consider the specificity of functional localization characterized by both the timing and the nature of the neural activation, for which internal physiological signals (for example, the plasma insulin) should be used as control. Temporal clustering is not only a new method for extracting the moment of the maximal response from a vast ensemble, but also a concept of brain parcellation that accounts for timing and connectivity^{12,15,29}. Moreover, by showing the correlation with a biochemical measurement, TCA-based fMRI may provide access to the neural effects induced by pharmacological as well as physiological interventions, particularly when we have no knowledge of the real timing of the interaction. □

Methods

Subjects and experiment protocols

Twenty-one human subjects participated (11 males, 10 females; 34 ± 3 yr; no family history of diabetes mellitus; 3 subjects excluded from the results because of head motion). All subjects had normal glucose tolerance and fasted (no food or beverages) for 12 h before the MRI study. The fMRI experiments were performed at the Research Imaging Center using a 2T Prestige MRI scanner (GE/Elscent). A mid-sagittal section (10 mm thick) was scanned using a T₂-weighted gradient-echo pulse sequence (echo time = 40 ms; repetition time = 60 ms; flip angle = 20°; image matrix = 200 × 200; field of view = 20 × 20 cm). We chose this sequence instead of EPI because EPI is subject to relatively larger distortion from the susceptibility effects, which may deteriorate the images in the hypothalamic region. Also high-resolution (1 mm) was required to focus on the hypothalamus.

For each subject, a total of 240 images were acquired for 48 min (Fig. 1), with an initial 10 min of pre-glucose (or water) acquisition (50 images) followed by 38 min of post-glucose acquisition (190 images). The subject ingested 75 g D-dextrose dissolved in 296 ml water (orange flavoured) through a per-oral rubber tube. The total time to ingest the glucose solution completely was 1.5 ± 0.5 min. Eight of our 21 subjects also underwent a repeat fMRI scan on a separate day using an equivalent amount of distilled water as stimulation, while all other procedures remained same. Blood samples were obtained at 15-min intervals through a catheter placed in an antecubital vein, starting 15 min before glucose ingestion. The plasma insulin concentration was analysed in duplicate by radioimmunoassay¹⁰.

Data analysis

Although head motion was restricted using a facial mask¹², the fMRI images may be subject to motion artefacts due to swallowing or pulsation. We employed a centre of mass method and navigator-echo techniques for the control of motion artefacts³⁰. The subjects with

image movement larger than half the voxel size were excluded from the results. Both functional and anatomical images were co-registered to a reference image in Talairach space. TCA was performed on the voxel-wise fMRI time series $S = S(n)$ obtained by a normalization procedure:

$$S(n) = [I(n) - I_b]/I_b \quad (n = 1, 2, 3, \dots, 240)$$

where $I(n)$ is the MRI signal intensity at the n th image and I_b is the mean of MRI signal intensities for the baseline. For each voxel, the maximum of the absolute values of $S(n)$ was identified at a certain image time. At each image time, the number of voxels that reached the maximum was counted to generate a histogram (Fig. 2).

On the basis of the concept of local maxima^{24–28}, a time dimension was added into the data space (x, y, z, t, S) to identify temporal maxima. Here, the time, t , is equivalent to n as described above, and x, y, z are the Talairach coordinates transformed from the image matrix (x was limited to -5 mm to $+5$ mm). For each subject, the same image matrix size, 175×175 (instead of the original 200×200), was used to cover the actual brain slice, and the first 10 images were excluded from the analysis. So for each subject, the total number of voxels was 30,625 and the total number of imaging time points was 230. If the temporal maxima were evenly distributed (as in an ideal situation without stimulation), the mean should be close to the ratio of $30,625/230$, or a baseline of 133 (Fig. 2). The following gaussian function was used to fit the distribution in Fig. 2d:

$$N_{\max}(t) = \frac{A}{\sqrt{2\pi}\sigma^2} e^{-\frac{(t-t_{\max})^2}{2\sigma^2}} + B$$

Four fitting parameters were obtained for each peak: A is the sum of all voxels in the distribution sampled, B is the baseline level as described above, σ is the standard deviation, and $t_{\max}$ is the peak time. Here, we were interested in $t_{\max}$, which was the moment of the maximal change (1.9 min for the first peak and 10.3 min for the second peak). On the basis of $t_{\max}$, the full width at half-maximum (FWHM) was calculated and the time window was defined as twice the FWHM centred on $t_{\max}$.

Unpaired t -tests were run individually to compare the mean of the fMRI signal during each time window with that during the baseline and the t -values were pooled from all the subjects and transformed to z -scores¹². The functional maps (colour coded) were overlaid on the corresponding T₁-weighted image acquired with the same field of view and image matrix sizes as the T₂-weighted images. The group analysis by cross-subject averaging ($n = 18$) of the statistical scores reduced random motion artefacts that might appear in the activation maps. Note that the statistical tests were applied to all the voxels in the image plane in case there were some voxels that had a submaximal response or multiple peaks during the 38-min period after glucose intake.

Received 28 March; accepted 2 May 2000.

- Rolls, E. T. & Treves, A. in *Neural Network and Brain Function* (eds Rolls, E. T. & Treves, A.) 17–22 (Oxford Univ. Press, Oxford/New York, 1998).
- Saper, C. B. in *The Human Nervous System* (ed. Paxinos, G.) 389–413 (Academic, New York, 1990).
- Benzo, C. A. The hypothalamus and blood glucose regulation. *Life Sci.* **32**, 2509–2515 (1983).
- Rohner-Jeanrenaud, F., Robbioni, E., Ionescu, E., Sauter, J. F. & Jeanrenaud, B. Central nervous system regulation of insulin secretion. *Adv. Metab. Disord.* **10**, 193–220 (1983).
- Oomura, Y. in *Handbook of the Hypothalamus* (eds Morgane, P. J. & Panksepp, J.) 557–620 (Dekker, New York, 1981).
- Grossman, S. P. Role of the hypothalamus in the regulation of food and water intake. *Psychol. Rev.* **82**, 200–224 (1975).
- Tataranni, P. A. et al. Neuroanatomical correlates of hunger and satiation in humans using positron emission tomography. *Proc. Natl Acad. Sci. USA* **96**, 4569–4574 (1999).
- Gautier, J.-F. et al. Regions of the human brain affected during a liquid-meal taste perception in the fasting state: a positron emission tomography study. *Am. J. Clin. Nutr.* **70**, 806–810 (1999).
- Denton, D. et al. Correlation of regional cerebral blood flow and change of plasma sodium concentration during genesis and satiation of thirst. *Proc. Natl Acad. Sci. USA* **96**, 2532–2537 (1999).
- Matsuda, M. et al. Altered hypothalamic function in response to glucose ingestion in obese humans. *Diabetes* **48**, 1801–1806 (1999).
- Woods, S. C., Seeley, R. J., Porte, D. J. & Schwartz, M. W. Signals that regulate food intake and energy homeostasis. *Science* **280**, 1378–1383 (1998).
- Liu, Y., Gao, J.-H., Liotti, M., Pu, Y. & Fox, P. T. Temporal dissociation of parallel processing in the human subcortical outputs. *Nature* **400**, 364–367 (1999).
- Friston, K. J., Jezzard, P. & Turner, R. Analysis of functional MRI time-series. *Hum. Brain Mapp.* **1**, 153–171 (1994).
- Rosen, B. R., Buckner, R. L. & Dale, A. M. Event-related functional MRI: past, present, and future. *Proc. Natl Acad. Sci. USA* **95**, 773–780 (1998).
- Ngan, S.-C. & Hu, X. Analysis of functional magnetic resonance imaging data using self-organizing mapping with spatial connectivity. *Magn. Reson. Med.* **41**, 939–946 (1999).
- Bendettini, P. A., Jesmanowicz, A., Wong, E. C. & Hyde, J. S. Processing strategies for time course data sets in functional MRI of the human brain. *Magn. Reson. Med.* **30**, 161–173 (1993).
- Xiong, J., Gao, J.-H., Lancaster, J. L. & Fox, P. T. Assessment and optimization of functional MRI analyses. *Hum. Brain Mapp.* **4**, 153–167 (1996).
- Zatorre, R. J., Jones-Gottman, M., Evans, A. C. & Meyer, E. Functional localization and lateralization of human olfactory cortex. *Nature* **360**, 339–340 (1992).
- Kinomura, S. et al. Functional anatomy of taste perception in the human brain studied with positron emission tomography. *Brain Res.* **659**, 263–266 (1994).
- Scott, T. R., Yan, J. & Rolls, E. T. Brain mechanisms of satiety and taste in macaques. *Neurobiology* **3**, 281–292 (1995).
- Reiman, E. M. The application of positron emission tomography to the study of normal and pathologic emotions. *J. Clin. Psychiatry* **58**, S4–12 (1997).
- Le Douarin, J. E. in *Handbook of Physiology* (eds Plum, F. & Mountcastle, V. B.) 419–459 (Am. Physiol. Soc., Washington, DC, 1987).

23. Liu, Y. *et al.* Temporal differentiation of the global effect and regional effect of microwave heating on the monkey cerebral circulation. *NeuroImage* **9**, S120 (1999).
24. Fox, P. T. *et al.* Mapping human visual cortex with positron emission tomography. *Nature* **323**, 806–809 (1986).
25. Fox, P. T., Mintun, M. A., Reiman, E. M. & Raichle, M. E. Enhanced detection of focal brain responses using intersubject averaging and change-distribution analysis of subtracted PET images. *J. Cereb. Blood Flow Metab.* **8**, 642–653 (1988).
26. Worsley, K. J., Evans, A. C., Marrett, S. & Neelin, P. A three dimensional statistical analysis for CBF activation studies in human brain. *J. Cereb. Blood Flow Metab.* **12**, 900–918 (1992).
27. Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C. & Evans, A. C. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping* **3**, 210–220 (1994).
28. Worsley, K. J. Local maxima and the expected Euler characteristic of excursion sets of χ^2 , F and t fields. *Adv. Appl. Prob.* **26**, 13–42 (1994).
29. Buechel, C., Coull, J. T. & Friston, K. J. The predictive value of changes in effective connectivity for human learning. *Science* **283**, 1538–1541 (1999).
30. Liu, H.-L. *et al.* Comparison of the center of mass and navigator echo corrections of image shift induced by central frequency shift in EPI fMRI. *Proc. Int. Soc. Magn. Reson. Med.* **7**, 271 (1999).

Acknowledgements

We thank M. Matsuda, R. A. DeFronzo, L. Nickerson, S. C. Pridgen and M. Liotti for their help in this study.

Correspondence and requests for materials should be addressed to Y.L. (e-mail: yijunliu@ufl.edu) or J.-H.G. (e-mail: gao@uthscsa.edu).

Postsynaptic translation affects the efficacy and morphology of neuromuscular junctions

Stephan J. Sigrist*, Philippe R. Thiel*, Dierk F. Reiff*, Pascal E. D. Lachance†, Paul Lasko† & Christoph M. Schuster*

* Friedrich-Miescher-Laboratorium der Max-Planck-Gesellschaft, Spemannstrasse 37-39, 72076 Tübingen, Germany

† Department of Biology, McGill University, 1205 Avenue Docteur Penfield, Montréal, Québec H3A 1B1, Canada

Long-term synaptic plasticity may be associated with structural rearrangements within the neuronal circuitry^{1,2}. Although the molecular mechanisms governing such activity-controlled morphological alterations are mostly elusive, polysomal accumulations at the base of developing dendritic spines³ and the activity-induced synthesis of synaptic components suggest that localized translation is involved during synaptic plasticity^{4,5}. Here we show that large aggregates of translational components as well as messenger RNA of the postsynaptic glutamate receptor subunit DGluR-IIA⁶ are localized within subsynaptic compartments of larval neuromuscular junctions of *Drosophila melanogaster*. Genetic models of junctional plasticity⁷ and genetic manipulations using the translation initiation factors eIF4E⁸ and poly(A)-binding protein⁹ showed an increased occurrence of subsynaptic translation aggregates. This was associated with a significant increase in the postsynaptic DGluR-IIA protein levels and a reduction in the junctional expression of the cell-adhesion molecule Fasciclin II. In addition, the efficacy of junctional neurotransmission and the size of larval neuromuscular junctions were significantly increased. Our results therefore provide evidence for a postsynaptic translational control of long-term junctional plasticity.

Translational control is primarily exerted by regulation of the initiation step of translation¹⁰, which appears to be controlled by the rate-limiting initiation factor eIF4E⁸. In addition, the interaction of the 5' cap bound eIF4E with the 3' end of mRNAs through a complex of other initiation factors and the poly(A)-binding protein

(PABP)⁹ has been shown to synergistically facilitate translation initiation¹¹. To assess the potential role of regulated translation during the development of the larval neuromuscular junctions (NMJs) in *Drosophila*, we analysed the subcellular expression pattern of eIF4E and PABP in filet preparations of third instar larvae. Both antigens showed a weak and ubiquitous expression in the cytoplasm of all larval cells, and they colocalized (Fig. 1a, inset) in strongly immunopositive aggregates up to 2 μ m in length (Fig. 1a, b, red panels) close to NMJs. Neuromuscular junctions have been highlighted by the immunodetection of the junctionally expressed cell-adhesion molecule Fasciclin II (FasII; ref. 12) (Fig. 1a, b, green panels). The specific localization of eIF4E/PABP aggregates close to and partially overlapping with junctional profiles (Fig. 1a, b, arrows and arrowheads) revealed that eIF4E/PABP aggregates are positioned subsynaptically within or adjacent to the subsynaptic reticulum (SSR). We did not find evidence for presynaptic or axonal localization of such aggregates. Therefore, the almost exclusive subsynaptic distribution of the eIF4E/PABP aggregates within larval muscles indicates that there may be a functional relationship

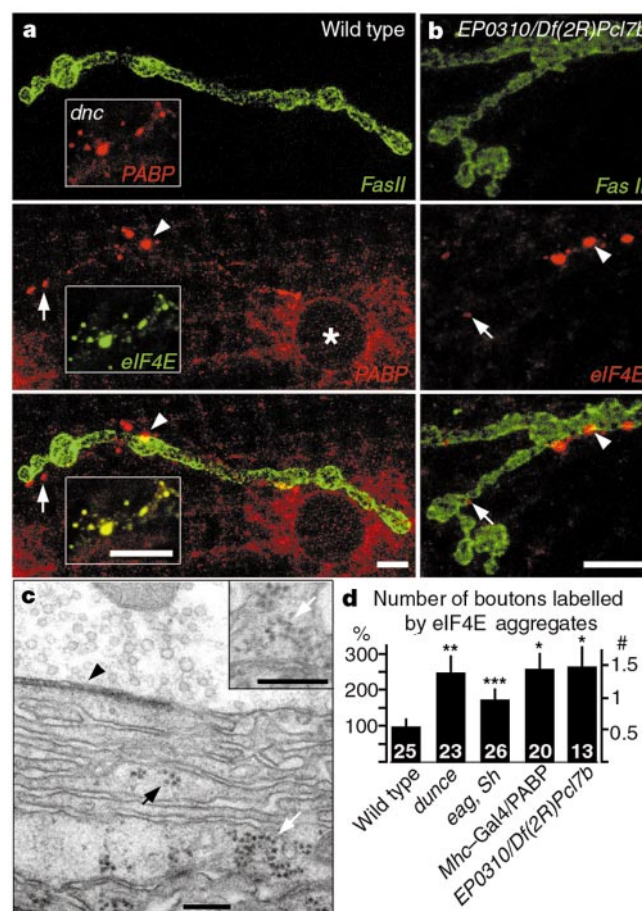


Figure 1 Subsynaptic translation aggregates and their regulation. Both PABP (a) and eIF4E (b) form strongly immunopositive aggregates of variable size (red channel), which are located close to (arrow) or overlapping (arrowhead) with FasII-labelled junctional profiles (green channel). Insets, PABP and eIF4E colocalize in junctional aggregates (yellow colour in bottom panel). Scale bars, 5 μ m. c, Ultrastructural examination of NMJs reveals that polysomes are localized within and close to the subsynaptic reticulum of type I boutons (arrows). Arrowhead marks an electron-dense area (synapse) with multiple presynaptic vesicles (above). Inset, circular polysomal profile within the SSR. Scale bars, 200 nm. d, The average number of boutons per NMJ (muscle 6/7, segments A2–5), which are labelled by subsynaptic translation aggregates (#), is significantly increased in animals representing both genetic gain-of-function (*Mhc-Gal4/PABP*) and loss-of-function (*EP0310/Df(2R)Pcl7b*) conditions of *pabp* (*t*-test: asterisk, $P < 0.0005$) and in the mutants *eag*, *Sh* and *dunce* (*t*-test: two asterisks, $P < 0.001$; three asterisks, $P < 0.005$). Data represent mean $\pm$ s.e.m. White numbers, number of animals analysed.

between NMJs and the appearance of nearby eIF4E/PABP aggregates.

Ultrastructural examinations of larval NMJs revealed polysomal accumulations within and close to the SSR (Fig. 1c, arrows). According to their variable size, subsynaptic location and frequency of detection, the larger of these polysomal clusters are likely to represent the eIF4E/PABP aggregates detected by light microscopy (Fig. 1c, white arrows). In addition, smaller polysomal aggregates were widely distributed in discrete membranous compartments throughout the SSR (Fig. 1c, black arrow), whereas presynaptic

and axonal profiles were free of polysomes. We therefore conclude that mRNAs are translated within subsynaptic compartments of larval NMJs (see Fig. 2d) and that local centres of concentrated, subsynaptic translation are identified by large junctional eIF4E/PABP aggregates.

To assess whether junctional translation is subject to regulation, we quantified the number of synaptic specializations (boutons) per NMJ that were labelled by one or more translation aggregates (Fig. 1d). Animals that overexpressed PABP in larval muscles and larvae that were mutant in *pabp* showed a significantly increased occurrence of subsynaptic eIF4E/PABP aggregates (Fig. 1d, right two bars) and an unaltered level of muscular PABP staining. In addition, the total PABP levels in crude larval protein extracts were unaltered in all analysed genotypes, even when PABP mRNA levels were significantly increased or reduced under genetic gain-of-function or loss-of-function conditions, respectively (Fig. 2g). Such a homeostasis of total PABP levels is a well described phenomenon for PABP¹³, and in crude protein extracts it might have masked the significant local increase in the number of PABP aggregates observable within subsynaptic compartments of NMJs (Fig. 1d). Although the exact reason for this increase in the occurrence of eIF4E/PABP aggregates is unknown, a local perturbation of PABP levels owing to a previously described overshooting compensation of the PABP-homeostasis mechanism¹³ might facilitate formation of subsynaptic translation aggregates.

A similar increase in the frequency of postsynaptic translation aggregates was also observed in two mutants representing well established genetic models of long-term synaptic plasticity in *Drosophila*^{7,12,14} (Fig. 1d), the hyperactive K⁺-channel mutant *eag*, *Sh* and the cAMP-phosphodiesterase mutant *dunce*. Thus, increased neuronal activity levels (in *eag*, *Sh*) as well as elevated cellular cAMP levels (in *dunce*) are capable of inducing subsynaptic translation aggregate formation. These findings are consistent with the hypothesis that synaptic activity can control synaptic translation^{15–17}.

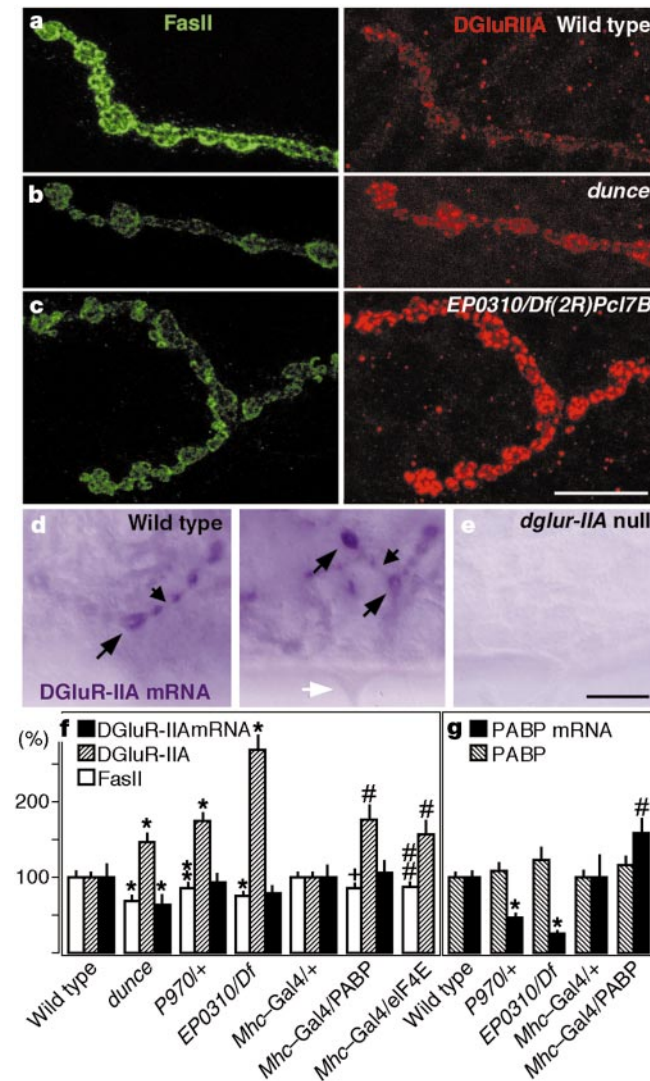


Figure 2 Targets of subsynaptic translation. **a–c, f**, Genotypes which increase the occurrence of subsynaptic translation aggregates (see Fig. 1d) show significantly reduced junctional FasII immunoreactivities (green channel in **a–c**, open bars in **f**) and strongly increased levels of synaptic DGlurIIA expression (red channel in **a–c**, hatched bars in **f**). The total DGlurIIA mRNA levels are unaltered in most analysed genotypes as quantified by RT–PCR (black bars in **f**), suggesting that the increased DGlurIIA expression is due to a post-transcriptional mechanism. *t*-tests, compared with wild type: asterisk, $P < 0.0001$; two asterisks, $P < 0.03$; compared with *Mhc-Gal4/+*: hash, $P < 0.0005$; two hashes, $P < 0.005$; '+', $P < 0.02$. **d**, DGlurIIA mRNA is detected by *in situ* hybridization in a characteristic pattern within subsynaptic compartments of NMJs (black arrows, see text). Nerve profiles are free of staining (white arrow). **e**, The transcriptional null mutant *dglur-IIA⁰⁹/Df(2L)chl4* is free of DGlurIIA *in situ* signals. Scale bars, 10 μ m. **g**, The total PABP protein content of crude larval extracts (hatched bars) is unaltered in genetic loss-of-function and gain-of-function *pabp* genotypes (as confirmed by quantitative RT–PCR for PABP mRNA, black bars; *t*-test: compared with wild type: asterisk, $P < 0.0001$; compared with *Mhc-Gal4/+*: hash, $P < 0.007$) revealing a strong homeostasis of general PABP levels *in vivo*. Data represent mean $\pm$ s.e.m.

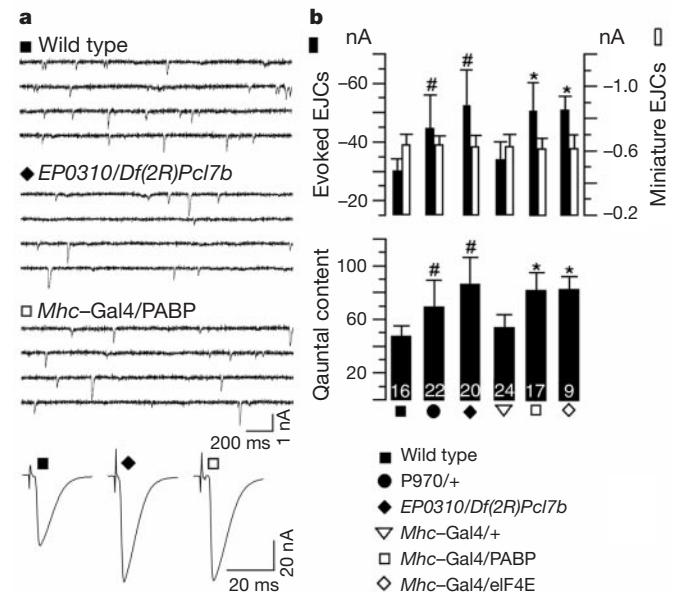


Figure 3 Subsynchronous translation affects junctional efficacy. **a**, Representative traces of miniature postsynaptic current recordings (mEJCs, upper panels) and average traces of ten consecutively recorded evoked EJCs (bottom panel) of the indicated genotypes. **b**, The mean amplitudes of the mEJCs are indistinguishable among all analysed genotypes (open bars). eEJCs are significantly increased in *pabp* mutants and in animals overexpressing PABP or eIF4E in muscles compared with wild type (hash) or control (asterisk, $P < 0.0001$). The derived quantal content and thus the junctional efficacy shows similarly significant increases in the analysed genotypes compared to controls. White numbers, number of analysed cells. Data represent mean $\pm$ s.e.m.

To identify potential substrates and targets of subsynaptic translation at larval NMJs, we performed quantitative immunostainings of several junctionally expressed proteins, including the synaptic vesicle protein synaptotagmin, the junctional anti-horseradish peroxidase (HRP) epitope, the cell-adhesion molecule Fasciclin II (FasII), the postsynaptic glutamate receptor subunit DGluR-IIA and the conventional myosin as a nonsynaptic protein. We did not detect obvious differences in the expression levels of myosin, synaptotagmin and the junctional anti-HRP immunoreactivity in all analysed genotypes (data not shown); however, animals that showed elevated numbers of subsynaptic translation aggregates (see Fig. 1d) consistently displayed increased junctional levels of DGluR-IIA (Fig. 2b, c, red panels) and an altered junctional distribution of FasII (Fig. 2b, c, green panels), which was associated with a significant reduction of synaptic FasII levels as compared with control animals (Fig. 2f, white bars). A similar FasII phenotype has been described in the plasticity models *eag*, *Sh* and *dunce*, and it has been shown that presynaptic FasII downregulation is essential for increased junctional outgrowth¹² (see below). Intriguingly, in *Aplysia* the FasII homologue apCAM is also presynaptically downregulated after treatments that increase synaptic efficacy and growth of new synaptic connections¹⁸. This synaptic apCAM regulation is thought to be achieved by a protein-synthesis-dependent activation of an endocytic apCAM internalization¹⁹. Given that FasII has been detected in membranes of a subset of presynaptic vesicles²⁰, it seems possible that subsynaptic protein synthesis affects junctional FasII levels through similar mechanisms to those in *Aplysia*.

The postsynaptic DGluR-IIA immunoreactivities were significantly stronger in translationally sensitized animals (Fig. 2b, c, red channels; Fig. 2f, hatched bars). This strong increase of synaptic DGluR-IIA expression was not due to transcriptional upregulation of *dglur-IIA*, as the total amounts of DGluR-IIA mRNAs were unaltered or even reduced in the analysed genotypes as compared

with controls (Fig. 2f, black bars). *In situ* hybridization experiments revealed that DGluR-IIA mRNA surrounds individual type-I boutons, with prominent staining of terminal and branch-point boutons (Fig. 2d, large arrows) and weak or absent staining within the SSR of interbouton connectives (Fig. 2d, small arrows). Thus, the subsynaptically localized DGluR-IIA mRNA represents a direct substrate for the junctional translation machinery. These results can not exclude an extrajunctional contribution to the observed synaptic DGluR-IIA increase, but they suggest that this phenotype is due to an increased subsynaptic synthesis of DGluR-IIA in genotypes with a higher occurrence of junctional eIF4E/PABP aggregates.

To analyse the functional consequences of genetically modified subsynaptic translation, we assessed the strength of neurotransmission at NMJs on muscle 6 of third instar larvae (Fig. 3). The average amplitudes of miniature excitatory junctional currents (mEJCs; Fig. 3a, top traces) and thus the quantal sizes were indistinguishable in all analysed genotypes (Fig. 3b, open bars). This finding indicates either that the additional receptor subunits that are synaptically localized (Fig. 2a–c, red panels) may be functionally silent (for example, through physiological silencing²¹ or intracellular localization²²) or that the amount of glutamate released from an individual quantum is not sufficient to saturate the postsynaptic receptors²³. In contrast, postsynaptic responses evoked by stimulation of motor nerve axons (Fig. 3a, bottom traces) were substantially larger in all mutants exhibiting increased levels of subsynaptic translation (Fig. 3b, top panel, filled bars). Thus, the derived quantal content was significantly increased above control values, suggesting that the observed larger amplitudes of evoked junctional responses arise from an increased number of released presynaptic vesicles per action potential.

To investigate whether the increase in junctional efficacy was due to a change in the number of synaptic specializations, we quantified the number of junctional boutons per NMJ (Fig. 4). Genotypes that displayed an increased occurrence of subsynaptic translation aggregates had significantly larger NMJs (Fig. 4c) and reduced junctional FasII levels (Fig. 2f). In addition, the junctional sizes of the analysed animals correlated in a highly significant manner with their estimated quantal contents (Fig. 4d), suggesting that junctional efficacy and the morphological elaboration of NMJs are tightly coupled. On the basis of our light microscopic examinations of DGluR-IIA labelled NMJs, the density of synapses within NMJs of all mutant animals appeared similar to that of controls or even higher, indicating that the total number of synapses increased proportionally with the junctional size. This finding indicates that the increased quantal content in animals with facilitated subsynaptic translation may be because of an increase in the number of vesicle release sites per given stimulus.

In summary, we have shown that translational machinery and mRNAs are associated with the subsynaptic reticulum of NMJs and that genetic manipulations that affect the occurrence of subsynaptic translation aggregates are accompanied by changes in the levels of synaptic proteins, such as DGluR-IIA and FasII. These same manipulations also affected the function and morphology of NMJs, suggesting that subsynaptic translation can instruct junctional growth and synaptic reorganization and thereby long-term functional changes. Our results further suggest that subsynaptic translation can be regulated by altered levels of neuronal activity, indicating that the regulation of postsynaptic translation participates in activity-dependent junctional plasticity. Thus, the inducible recruitment of postsynaptic protein synthesis appears to render individual synapses competent to instruct long-term changes in their functions and morphological organization. Given that localized protein synthesis has been shown to act in a synapse specific stabilization of long-term facilitation in central neurons of *Aplysia*^{24,25}, it emerges that synaptic translation might represent a common principle of long-term alterations of neuronal function and connectivity.

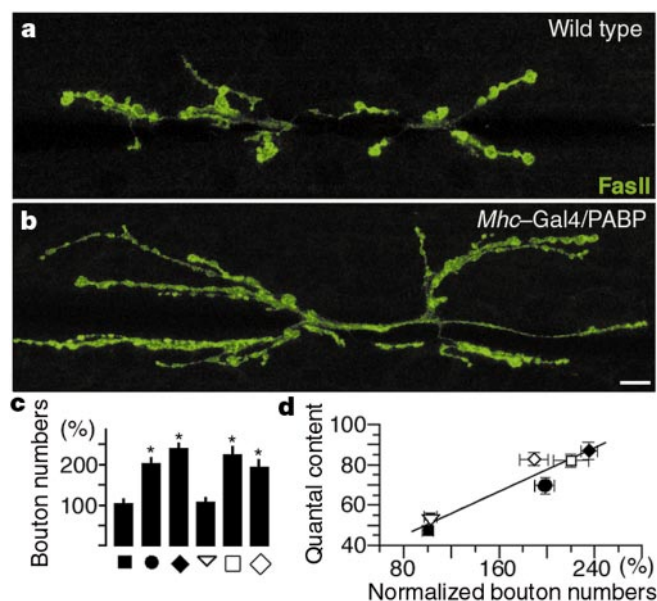


Figure 4 Subsynaptic translation instructs morphological and functional growth of NMJs. **a–c**, Genetic manipulations that increase the occurrence of subsynaptic translation aggregates, either through a series of *pabp* alleles (*pabp⁹⁷⁰/+* ($n = 50$), filled circles; *pabp^{EP0310}/Df(2R)Pcl7b* ($n = 37$), filled diamonds) or through the targeted overexpression of PABP (**b**) or eIF4E in larval muscles (*Mhc-Gal4/PABP* ($n = 20$), open squares; *Mhc-Gal4/eIF4E* ($n = 10$), open circles) result in significant size increases of larval NMJs compared with controls (wild type ($n = 36$), filled squares; *Mhc-Gal4/+* ($n = 27$), open triangles; *t*-test: asterisk, $P < 0.0001$). Data in **c** are plotted as means $\pm$ s.e.m. Scale bar, 10 μ m. **d**, The quantal content of the analysed genotypes is plotted as a function of junctional size. A linear regression analysis of the data revealed a highly significant correlation between junctional size and junctional efficacy ($r = 0.955$; $P < 0.001$).

Methods

Molecular genetics

Df(2R)Pcl7B removes the *pabp* coding region and 3' untranslated region (UTR). *pabp*^{k10109} (*P970*) is a P-element insertion into the exon preceding the first coding exon of *pabp*. *P970/Df(2L)Pcl7B* is embryonically lethal; a precise excision of *P970* restored viability and reverted the morphological phenotypes of *P970/+* close to wild type, whereas an imprecise excision of *P970*, which removed the transcription start site of *pabp*, phenocopied *P970*. The P-element *EP0310* is inserted into the first 3' non-coding exon of *pabp*. We also generated an UAS-*pabp* transgene containing the PABP coding region without 5' and 3' UTRs inserted into pUAST for Gal4-controlled expression. The UAS-*elF4E* transgenic flies were provided by R. Rivera-Pomar. The *Mhc*-Gal4 and *elav*-Gal4 lines allowed the expression of UAS transgenes in all muscles or all neurons, respectively. Mutations in the *dglur-IIA* locus (*dglur-IIA*⁸⁹ and *Df(2L)clh4*) have been described²⁶. To provide a control for the junctional effects of generally reduced translation, we examined four *Minute* mutants (*M(2)58F*, *M(2)36F*, *M(2)53-1*, *M(2)24F-1*), which are defective in constitutive components of ribosomal subunits²⁷. These *Minute* mutants caused a significant developmental delay and somewhat smaller larval body sizes, but no obvious molecular or morphological junctional phenotypes.

Quantitative polymerase chain reaction with reverse transcription

Total RNA was extracted from 20 to 70 early to mid-third instar larvae. Two independent RNA preparations per genotype were transcribed and each cDNA was subjected to multiplex PCR²⁸ 6 to 15 times. As an internal standard we used oligonucleotides specific for glyceraldehyde-3-phosphate dehydrogenase (GAPDH), whose invariant expression in all analysed genotypes was confirmed by comparison to the myosin heavy chain mRNA level. The amount of PABP- or DGluR-IIA-specific PCR product was quantified and normalized to the GAPDH-specific PCR product.

Antibodies

The rabbit anti-PABP antiserum was raised against the peptide Leu535–Lys552, affinity purified and specificity controlled by peptide competition experiments and comparison to a previously characterized anti-PABP serum (R. Rivera-Pomar). The rabbit anti-eIF4E antiserum was raised against a bacterially expressed glutathione-S-transferase (GST)–eIF4E fusion protein. The affinity-purified antiserum detects the two isoforms of eIF4E²⁹ on western blots. Antibodies were gifts of C. S. Goodman (FasII, 1D4; myosin, FMM5), J. Kidokoro (DGluR-IIA, DM2) and T. Littleton (synaptotagmin). The anti HRP-antibody recognizes a neural epitope in insects³⁰ and was purchased from Sigma.

Immunofluorescence quantification

All larvae used in this study were raised under normalized culture conditions (25 °C, 65% humidity, high animal density). Mid-third instar larvae of similar age and body size were processed for immunofluorescent detection as described¹². Junctional immunoreactivity levels of DGluR-IIA and FasII were quantified in triple-labelled larval preparations with the invariant anti-HRP immunoreactivity at NMJs as an internal staining standard. Five to nine type Ib boutons (muscle 6/7, abdominal segment 2) were selected in the anti-HRP channel of a recorded confocal image stack and the average fluorescence signal of this selection was determined for all three channels. The signal ratios DGluR-IIA/HRP and FasII/HRP of at least 2 non-overlapping areas per NMJ were accumulated from 14–30 animals per genotype.

PABP–protein quantification

Twenty male larvae of the indicated genotypes were homogenized, and the crude protein extract equivalent of two animals was immunoblotted and probed with affinity-purified anti-PABP serum and anti-tubulin antibody (Amersham). The anti-PABP immunoreactivity was quantified and normalized to the anti-tubulin reaction on the basis of three to four independent extracts.

Junctional size quantification

As none of the genotypes used here showed systematic alterations of larval muscle sizes (average of all scored muscles of all genotypes: 47 ± 8.5 square scale units, $n = 333$), we used the measured muscle surface area as a fine-scale staging criterion to normalize the bouton counts per muscle 6/7 NMJ (abdominal segment 2).

Electrophysiology

Third instar larvae were dissected and prepared for intracellular recordings as described¹². Miniature and evoked postsynaptic currents were recorded from muscle fibre 6 of abdominal segments 2 and 3 in TEVC mode (Axoclamp 2B, Axon Instruments). For stimulation, the cut end of the intersegmental nerve was placed into a suction electrode and suprathreshold current pulses were applied at 0.1 Hz. For recordings, muscle cells were impaled with two 15–30 M Ω microelectrodes filled with 2 M KCl (the resistance of the current passing electrode was usually 5–10 M Ω lower than that of the voltage sensing electrode). Cells with a resting potential of less than –60 mV in HL3 solution (1 mM Ca²⁺) were selected for further analysis. Clamp settling times in response to voltage steps from –60 to –70 mV were 300–600 μ s and voltage errors were up to 4 mV when eEJCs were close to 100 nA. eEJCs (VC at –60 mV) were low-pass filtered at 2 kHz, mEJCs (voltage clamp at –70 mV) at 500 Hz and subsequently digitized. Thirty to fifty eEJCs and 90 s of mEJCs recordings were used per cell for off-line analysis (pClamp6, Axon Instruments;

Jaen Software, Leonia). Estimates of the quantal content were derived from these data by dividing the mean eEJC through the mean mEJC of each analysed cell.

Received 22 February; accepted 11 April 2000.

- Maletic-Savatic, M., Malinow, R. & Svoboda, K. Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity. *Science* **283**, 1923–1927 (1999).
- Engert, F. & Bonhoeffer, T. Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature* **399**, 66–70 (1999).
- Stewart, O. & Falk, P. M. Protein-synthetic machinery at postsynaptic sites during synaptogenesis: a quantitative study of the association between polyribosomes and developing synapses. *J. Neurosci.* **6**, 412–423 (1986).
- Schuman, E. M. mRNA trafficking and local protein synthesis at the synapse. *Neuron* **23**, 645–648 (1999).
- Kiebler, M. A. & DesGroseillers, L. Molecular insights into mRNA transport and local translation in the mammalian nervous system. *Neuron* **25**, 19–28 (2000).
- Schuster, C. M. *et al.* Molecular cloning of an invertebrate glutamate receptor subunit expressed in *Drosophila* muscle. *Science* **254**, 112–114 (1991).
- Budnik, V., Zhong, Y. & Wu, C.-F. Morphological plasticity of motor axon terminals in *Drosophila* mutants with altered excitability. *J. Neurosci.* **10**, 3754–3768 (1990).
- Sonenberg, N. & Gingras, A.-C. The mRNA 5' cap-binding protein eIF4E and control of cell growth. *Curr. Opin. Cell Biol.* **10**, 268–275 (1998).
- Gallie, D. R. A tale of two termini: A functional interaction between the termini of a mRNA is a prerequisite for efficient translation initiation. *Gene* **216**, 1–11 (1998).
- Mathews, M. B., Sonenberg, N. & Hershey, J. W. B. in *Translational Control* (eds Hershey, J. W. B., Mathews, M. B. & Sonenberg, N.) 1–29 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1996).
- Craig, A. W. B., Haghighat, A., Yu, A. T. K. & Sonenberg, N. Interaction of polyadenylate-binding protein with the eIF4G homologue PAIP enhances translation. *Nature* **392**, 520–523 (1998).
- Schuster, C. M., Davis, G. W., Fetter, R. D. & Goodman, C. S. Genetic dissection of structural and functional components of synaptic plasticity. II. Fasciclin II controls presynaptic structural plasticity. *Neuron* **17**, 655–667 (1996).
- Wu, J. & Bag, J. Negative control of the poly(A)-binding protein mRNA translation is mediated by the adenine-rich region of its 5'-untranslated region. *J. Biol. Chem.* **273**, 34535–34542 (1998).
- Balling, A., Technau, G. M. & Heisenberg, M. Are the structural changes in adult *Drosophila* mushroom bodies memory traces? Studies on biochemical learning mutants. *J. Neurogenet.* **4**, 65–73 (1987).
- Weiler, I. J. *et al.* Fragile X Mental Retardation Protein is translated near synapses in response to neurotransmitter activation. *Proc. Natl Acad. Sci. USA* **94**, 5395–5400 (1997).
- Wu, L. *et al.* CPEB-mediated cytoplasmic polyadenylation and the regulation of experience-dependent translation of a-CaMKII mRNA at synapses. *Neuron* **21**, 1129–1139 (1998).
- Nayak, A., Zastrow, D. J., Lickteig, R., Zahniser, N. R. & Browning, M. D. Maintenance of late-phase LTP is accompanied by PKA-dependent increase in AMPA receptor synthesis. *Nature* **394**, 680–683 (1998).
- Mayford, M., Barzilai, A., Keller, F., Schacher, S. & Kandel, E. R. Modulation of an NCAM-related adhesion molecule with long-term synaptic plasticity in *Aplysia*. *Science* **256**, 638–644 (1992).
- Bailey, C. H., Chen, M., Keller, F. & Kandel, E. R. Serotonin-mediated endocytosis of apCAM: an early step of learning-related synaptic growth in *Aplysia*. *Science* **256**, 645–649 (1992).
- Schuster, C. M., Davis, G. W., Fetter, R. D. & Goodman, C. S. Genetic dissection of structural and functional components of synaptic plasticity. I. Fasciclin II controls synaptic stabilization and growth. *Neuron* **17**, 641–654 (1996).
- Davis, G. W., DiAntonio, A., Petersen, S. A. & Goodman, C. S. Postsynaptic PKA controls quantal size and reveals a retrograde signal that regulates presynaptic transmitter release in *Drosophila*. *Neuron* **20**, 305–315 (1998).
- Shi, S. H. *et al.* Rapid spine delivery and redistribution of AMPA receptors after synaptic NMDA receptor activation. *Science* **284**, 1811–1816 (1999).
- Liu, G. S., Choi, S. W. & Tsien, R. W. Variability of neurotransmitter concentration and nonsaturation of postsynaptic AMPA receptors at synapses in hippocampal cultures and slices. *Neuron* **22**, 395–409 (1999).
- Casadio, A. *et al.* A transient, neuron-wide form of CREB-mediated long-term facilitation can be stabilised at specific synapses by local protein synthesis. *Cell* **99**, 221–237 (1999).
- Sherff, C. M. & Carew, T. J. Coincident induction of long-term in *Aplysia*: Cooperativity between cell bodies and remote synapses. *Science* **285**, 1911–1914 (1999).
- Petersen, S. A., Fetter, R. D., Noordermeer, J. N., Goodman, C. S. & DiAntonio, A. Genetic analysis of glutamate receptors in *Drosophila* reveals a retrograde signal regulating presynaptic transmitter release. *Neuron* **19**, 1237–1248 (1997).
- Berkeley *Drosophila* Genome Project.
- Steinbach, O. C. & Rupp, R. A. W. in *Methods in Molecular Biology*. Vol. 127. *Molecular Methods in Developmental Biology: Xenopus and Zebrafish* (ed. Guille, M.) 41–56 (Humana, Totowa, New Jersey, 1999).
- Lavoie, C. A., Lachance, P. E. D., Sonenberg, N. & Lasko, P. Alternatively spliced transcripts from the *Drosophila* eIF4E gene produce two different cap-binding proteins. *J. Biol. Chem.* **271**, 16393–16398 (1996).
- Gorczyca, M., Augart, C. & Budnik, V. Insulin-like receptor and insulin-like peptide are localized at neuromuscular junctions in *Drosophila*. *J. Neurosci.* **13**, 3692–3704 (1993).

Acknowledgements

We are indebted to C. S. Goodman, J. Kidokoro, T. Littleton and R. Rivera-Pomar for their gifts of reagents. We thank E. M. Illgen and M. Langege for technical assistance, and we are grateful to H. Schwarz for his support and advice during ultrastructural studies. We would like to thank W. Hoch and J. Knoblich for comments on the manuscript and helpful discussions. P.E.D.L. and P.L. are grateful to the MRC of Canada for financial support.

Correspondence and requests for material should be addressed to C.M.S. (e-mail: Christoph.Schuster@tuebingen.mpg.de).

Production of gene-targeted sheep by nuclear transfer from cultured somatic cells

K. J. McCreath, J. Howcroft, K. H. S. Campbell*, A. Colman, A. E. Schnieke & A. J. Kind

PPL Therapeutics Ltd, Roslin, Edinburgh, EH25 9PP, UK

* Present address: University of Nottingham, School of Biological Sciences, Division of Animal Physiology, Loughborough, LE12 5RD, UK

It is over a decade since the first demonstration that mouse embryonic stem cells could be used to transfer a predetermined genetic modification to a whole animal¹. The extension of this technique to other mammalian species, particularly livestock, might bring numerous biomedical benefits, for example, ablation of xenoreactive transplantation antigens, inactivation of genes responsible for neuropathogenic disease and precise placement of transgenes designed to produce proteins for human therapy. Gene targeting has not yet been achieved in mammals other than mice, however, because functional embryonic stem cells have not been derived. Nuclear transfer from cultured somatic cells provides an alternative means of cell-mediated transgenesis^{2,3}. Here we describe efficient and reproducible gene targeting in fetal fibroblasts to place a therapeutic transgene at the ovine $\alpha 1(I)$ procollagen (*COL1A1*) locus and the production of live sheep by nuclear transfer.

We previously showed that transfection of fetal fibroblasts with nuclear transfer offers an efficient and practical method of producing sheep carrying randomly integrated transgenes³, and similar work has been reported in cattle⁴. Gene targeting is a more powerful method of genetic manipulation and requires essentially the same procedures of transfection and drug selection of cultured cells. Although experimental gene targeting was first carried out in

somatic cell lines^{5,6}, the use of embryonic stem (ES) cells now predominates; however, there has been no definitive comparison of gene targeting efficiency in primary somatic cells and ES cells (but see ref. 7 for review). We wished to determine whether practically useful gene targeting could be achieved in primary ovine fetal fibroblasts, and whether these cells could produce viable animals by nuclear transfer. The ovine *COL1A1* gene represented a suitable target with which to establish gene targeting in fetal fibroblasts for three reasons. First, we expected that gene-targeting events would be very rare compared with random integrations. *COL1A1* is highly expressed in fibroblasts, allowing promoter-trap enrichment of gene-targeting events. Second, few ovine genes have been cloned and characterized. *COL1A1* is well studied and highly conserved in several species, facilitating molecular cloning and construction of ovine gene-targeting vectors. Third, mutations in *COL1A1* can cause connective tissue disorders in humans, for example, *osteogenesis imperfecta*^{8,9}. The ability to generate models of human genetic disorders by gene targeting in animals other than mice might be valuable for clinical research; however, we chose to target a site that would not significantly affect type I collagen protein function or expression to avoid affecting fetal development.

We used two gene-targeting vectors to target ovine *COL1A1* (Fig. 1). Each vector incorporated two regions of *COL1A1* homology, derived from a contiguous fragment of Poll Dorset fetal fibroblast³ (PDFF2) genomic DNA. COLT-1 was designed to insert a promoterless *neo* selectable marker between the *COL1A1* translational stop and polyadenylation signal, such that transcription of the targeted locus resulted in a bicistronic messenger RNA. An internal ribosomal entry site (IRES)¹⁰ immediately 5' of *neo* facilitated translation. COLT-2 had the same structure, but included a transgene as a separate transcription unit located 3' of *neo*. This transgene, termed AATC2, comprised human $\alpha 1$ -antitrypsin (AAT) complementary DNA within an ovine β -lactoglobulin (BLG) expression vector designed to direct expression in the lactating mammary gland¹¹.

We transfected COLT-1 DNA into early passage PDFF2 female and PDFF5 male ovine primary fetal fibroblasts; we transfected COLT-2 DNA into PDFF2 cells. Stable G418 resistant clones

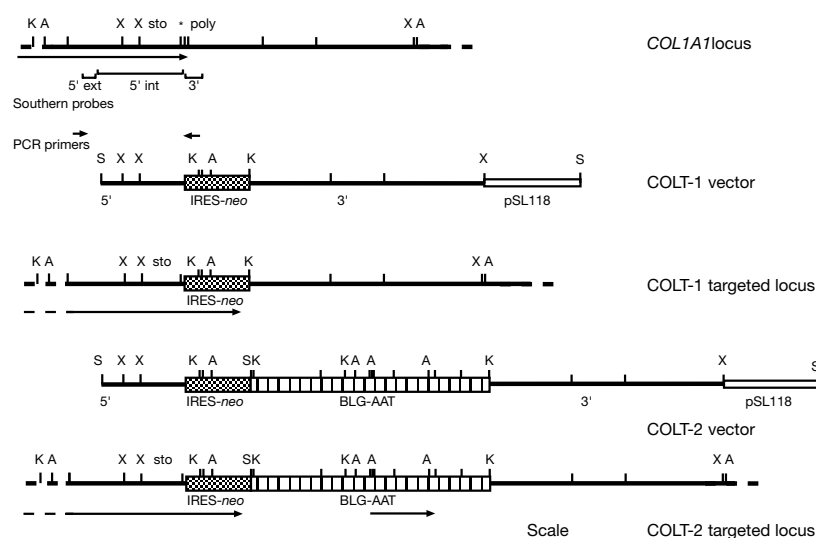


Figure 1 Diagrams of ovine *COL1A1*, COLT-1 and COLT-2 targeting vectors and targeted *COL1A1* locus. The map of ovine *COL1A1* shows the translational stop and polyadenylation sites, the direction of transcription and extent of the *COL1A1* mRNA is indicated by an arrow, an asterisk marks an *Ssp1* site where the targeted gene insertions were made. Southern hybridization probes and PCR primers are indicated. The maps of the COLT-1 and COLT-2 vectors and the targeted *COL1A1* locus show the IRES-*neo*

cassette as a shaded box, the BLG driven AAT transgene by a striped box, and the bacterial vector pSL1180 by an open box. The direction and predicted extent of the *COL1A1* / IRES-*neo* bicistronic mRNA and the AAT transgene mRNA are shown as arrows. The scale bar represents 2 kb. Restriction enzyme sites: A, *Asp1*; B, *Bam*HI; Sc, *Sac*II; S, *Sal*I; K, *Kpn*I; X, *Xho*I.

were derived. About 30 days of culture elapsed between fetal disaggregation and cryopreservation of gene-targeted cell clones. DNA samples of each cell clone were initially screened by polymerase chain reaction (PCR) using primers designed to amplify a 3.4-kilobase (kb) fragment across the 5' junction of the targeted locus (Fig. 1). In each case, a high proportion of G418 resistant cell clones were found to have undergone gene targeting (5 out of 36 PDFF2 COLT-1, 4 out of 56 PDFF5 COLT-1, and 46 out of 70 PDFF2 COLT-2 cell clones analysed). We called COLT-1 cell clones 'PDCOL' and COLT-2 cell clones 'PDCAAT'. The DNA sequence of PCR products amplified from three PDCOL and two PDCAAT cell clones was determined across the 5' junction; each was consistent with gene targeting (data not shown).

Figure 2a shows Southern analysis of the 5' junctions of a series of PDCAAT cell clones. All samples showed the presence of a 7-kb *Bam*HI fragment from the normal *COL1A1* locus. Each clone identified as positive by PCR also showed a diagnostic 4.7-kb *Bam*HI fragment spanning the 5' junction of the COLT-2 targeted locus. This is consistent with the presence of one targeted and one normal *COL1A1* allele. PDCAAT cell clones 87 and 99 also showed the presence of additional bands, indicating additional integrations of COLT-2.

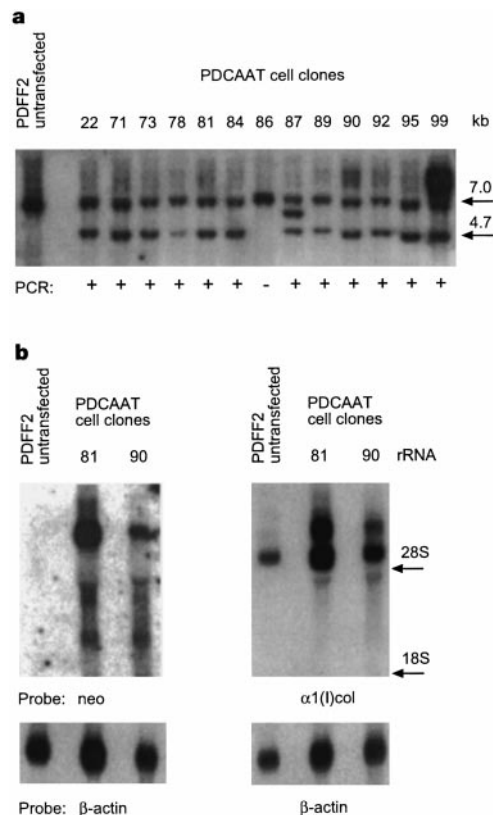


Figure 2 Analysis of COLT-2 transfected (PDCAAT) cell clones. **a**, Southern analysis. Each lane contains *Bam*HI digested genomic DNA from cell samples hybridized with a 3-kb *COL1A1* *Sal*, *Ssp*I fragment corresponding to the 5' homologous arm of COLT-1 and COLT-2 (see Fig. 1). The cell clone is indicated above each lane and results of the PCR screen are shown below each lane. The positions of the 7-kb *Bam*HI fragment from the non-targeted ovine *COL1A1* locus and the 4.7-kb *Bam*HI fragment from the COLT-2 targeted locus are marked. **b**, Northern analysis. Each lane contains 10 μ g total RNA extracted from PDFF2 cells and targeted cell clones PDCAAT81 and PDCAAT90 as indicated. Duplicate blots were hybridized with a *neo* probe (upper left), and full-length human $\alpha 1(I)$ procollagen cDNA (upper right). The positions of the 28S and 18S rRNA bands are indicated. The lower two autoradiographs show rehybridization of the same blots with mouse β -actin cDNA as an indication of the amount of total RNA loaded.

PDFF5 cells were also targeted at high frequency with COLT-1. PDFF2 and PDFF5 cells have different parentage from within the PPL outbred flock of Poll Dorset sheep. This indicates that it may not always be essential to isolate DNA from the same individual, or an animal of the same inbred strain, to achieve efficient gene targeting¹²; however, the degree of sequence divergence, if any, between the targeted *COL1A1* alleles in these cells has not yet been determined.

Northern analysis of cell clones PDCAAT 81 and 90 is shown in Fig. 2b. Hybridization with human $\alpha 1(I)$ procollagen cDNA detected a 4.8-kb mRNA species in both non-transfected cells and PDCAAT cell clones, consistent with expression of normal *COL1A1*. A larger species of about 6.8 kb was present only in the PDCAAT cell clones, consistent with a bicistronic *COL1A1*-IRES-*neo* fusion mRNA. Hybridization of the same RNA samples with a *neo* probe also detected a 6.8-kb mRNA in the targeted clones, again consistent with a bicistronic mRNA. These results confirmed that gene targeting had occurred. This analysis also showed that, unlike mouse, rat and human, which express two endogenous $\alpha 1(I)$ procollagen mRNA species from different polyadenylation sites^{13,14}, sheep express a single mRNA species.

Although these experiments were designed to avoid disruption of *COL1A1* gene expression, the mRNA from the targeted locus is less abundant than the wild type (Fig. 2b). Whether this reflects different mRNA stability or transcriptional activity has yet to be determined. However, elements which affect transcription have been identified at the 3' end of *COL1A1* in other species^{15,16}, and targeted DNA insertion may affect their function.

We carried out northern analysis to determine whether placement of the AATC2 transgene adjacent to the highly expressed *COL1A1* gene resulted in aberrant expression of the BLG promoter in fibroblasts. Hybridization with human AAT cDNA failed to detect AAT mRNA expression in either PDCAAT81 or 90 cells (data not shown), indicating no apparent loss of BLG promoter specificity.

Four targeted cell clones, all derived from PDFF2 female cells (PDCOL6, PDCOL13, PDCAAT81 and PDCAAT90), were selected for nuclear transfer on the basis of their vigour and normal metaphase chromosome number. Nuclear transfer was carried out on twelve occasions and the results are summarized in Table 1. Fourteen lambs were live-born: seven died within 30 hours of birth, and one each after 3 days, 8 days, 7.5 weeks and 12 weeks. Three lambs are currently alive and thriving at almost one year of age. The first two live-born lambs are shown in Fig. 3.

Post mortem examination of lambs that died *in utero* or after birth revealed a range of abnormalities. Although there was no consistent pattern, we observed a high incidence of kidney defects (frequently renal pelvis dilation), liver and brain pathology. These findings are similar to a previous nuclear transfer study using the same cells³,

Table 1 Summary of nuclear transfer results

Nuclear donor cells	PDCOL 6	PDCOL 13	PDCAAT 81	PDCAAT 90
Reconstructed embryos	109	154	71	83
Embryos recovered from temporary recipient	104	149	62	78
Embryos developed to morula or blastocyst	14	43	4	19
Embryos transferred to final recipients	14	43	4	19
Final recipients	8	22	2	10
Fetuses at day 60	5*	10*	2	3*
Live-born lambs	4	8	0	2
Lambs alive beyond 1 week	2	3	0	1
Lambs alive beyond 6 months (lamb ID)	1	1	0	1
	(ID 990502)	(ID 990504)		(ID 990507)

*One twin pregnancy.

and are therefore probably due to some aspect of the cell treatment or nuclear transfer procedure and not a consequence of gene targeting *per se*. Several researchers have reported developmental abnormalities associated with somatic cell nuclear transfer^{4,17,18}. Although there is some indication that inappropriate expression of imprinted genes may be involved¹⁹, definitive investigations have yet to be carried out. Understanding and rectifying this problem is a continuing priority.

Figure 4 shows Southern analysis of 5' and 3' junctions of the *COL1A1* targeted locus in representative nuclear transfer lambs. Hybridization with the same 5' and 3' probes used to analyse PDCAAT cell clones revealed fragments consistent with the presence of one targeted and one normal *COL1A1* allele in the two lambs shown. This was confirmed using a 5' probe external to the vector.

Sixteen lambs and fetuses were analysed by Southern blot, of which fifteen confirmed the presence of a targeted allele. One lamb (990504), derived from PDCOL13, showed only the normal *COL1A1* gene, which probably indicates that the cell isolate was oligoclonal. Non-transfected cells have been detected within some G418 selected cell isolates in previous experiments (unpublished data).

Lamb 990507 derived from clone PDCAAT90 was hormonally induced to lactate and milk samples analysed by western blotting (data not shown). AAT was detected at a concentration of

650 $\mu\text{g ml}^{-1}$, which compares favourably with the highest level previously reported for an AAT cDNA transgene in sheep carrying multiple random gene inserts (18 $\mu\text{g ml}^{-1}$)²⁰. This indicates that the *COL1A1* locus supports transgene expression even though it is not actively expressed in mammary epithelium²¹.

We have shown that gene targeting can be carried out efficiently in somatic cells and that viable animals can be produced by nuclear transfer. We have also obtained preliminary data (that is, PCR fragment size and sequence) indicating similarly efficient targeting at the α -1,3-galactosyl-transferase locus in porcine fibroblasts (unpublished data). Notably, the use of nuclear transfer does not require embryonic stem or embryonic germ cells, and circumvents the generation of chimaeric animals, which would be costly and time consuming in livestock. Fibroblasts are also being used in clinical trials to provide a protein production system after *ex vivo* gene therapy in human patients²², and it has been suggested that the introduction of therapeutic transgenes by homologous recombination could avoid undesirable effects arising from random integration²³. Nuclear transfer in animals such as sheep provides a rigorous means of testing the suitability of specific loci for transgene placement. If the *COL1A1*-targeted sheep continue to show no locus-related deleterious effects, this would indicate that this target locus may provide a permissive and benign environment for the insertion of therapeutically useful genes. □

Methods

Gene-targeting vectors

The promoter trap vector COLT-1 comprised a 3-kb region of the 3' end of the ovine *COL1A1* gene from a point roughly 2.9-kb 5' of the translation stop site to an *Spl* site 131-bp 3' of the stop site; a 0.6-kb IRES region¹⁰ corresponding to bases 1,247–1,856 of the pIREShyg vector (Clontech); a 1.7-kb region containing the bacterial *neomycin* gene and a portion of the 3' end of the human growth hormone gene containing the polyadenylation site, essentially as described²⁴; an 8.3-kb region of the 3' end and flanking region of the ovine *COL1A1* gene from an *Spl* site 131-bp 3' of the translational stop site to a *XhoI* site roughly 8.4-kb 3' of the stop site; and the bacterial cloning vector pSL1180 (Pharmacia). DNA fragments homologous to the ovine *COL1A1* gene were derived from a single genomic clone isolated from a library of genomic fragments of PDFF2 in bacteriophage λ . The promoter trap transgene placement vector COLT-2 was constructed by inserting an *MluI* fragment containing the AATC2 transgene into COLT-1 at a unique *EcoRV* site at the 3' end of the IRES–*neo* region.

Preparation, culture and transfection of primary fibroblasts

Derivation of ovine PDFF2 and PDFF5 cells has been described³. PDFF cells were grown throughout in BHK 21 (Glasgow MEM) medium supplemented with 2 mM glutamine, 1 mM sodium pyruvate, 1X non-essential amino acids and 10% fetal calf serum in standard tissue-culture vessels in a humidified atmosphere composed of 2% CO_2 , 5% O_2 and 93% N_2 , and were passaged by standard trypsinization. PDFF cells were used at passage three, and had undergone 5–6 days of culture following fetal disaggregation. Cells were plated at 5×10^5 cells in a 25-cm² flask and transfected the next day with either 6 μg of *Sall* linearized COLT-1 or *SacI* linearized COLT-2, using lipofectAMINE (Gibco, BRL Life Technologies) according to the manufacturer's guidelines. G418 selection (0.8 mg ml^{-1}) was applied 48 h after transfection. On average ~200 G418^r colonies were derived per 5×10^5 cells transfected. Well separated G418^r colonies were isolated, expanded and cryopreserved by standard procedures.

Nuclear transfer

Ovine cell clones were prepared for nuclear transfer by culture in medium containing reduced (0.5%) serum for either 2 or 4 days. Transfer of cell nuclei into Poll Dorset oocytes was carried out essentially as described³.

Nucleic acid analysis

Putative targeted cell clones were screened by PCR amplification across the 5' short arm of homology. The position of PCR primers is shown in Fig. 1. Samples of cell clones for screening were lysed in PCR lysis buffer (50 mM KCl, 1.5 mM MgCl_2 , 10 mM Tris-HCl pH 8.5, 0.5% Nonidet P40, 0.5% Tween, 400 $\mu\text{g ml}^{-1}$ Proteinase K) at 65 °C for 30 min. Proteinase K was inactivated at 95 °C for 10 min, and PCR amplification was performed using the 'Expand long template PCR system' (Boehringer) according to the manufacturer's recommended conditions. PCR primer sequences were: COLTPCR4 primer, 5'-GGTTTGTTCCTCCAGGTGCTCA-3'; COLTPCR8 primer, 5'-GACCTTG-CATTCCTTTG GCGAGAG-3'. Thermal cycling conditions were: 94 °C, 2 min; 10 cycles of 94 °C, 10 s, 55 °C, 30 s, 68 °C, 2 min; 20 cycles of 94 °C, 10 s, 60 °C, 30 s, 68 °C, 2 min + 20 s per cycle; followed by 68 °C, 7 min. PCR products were analysed by agarose gel electrophoresis.



Figure 3 Gene-targeted lambs. Lambs 990502 and 990503, both derived by transfer of nuclei from gene-targeted cell clone PDCOL6.

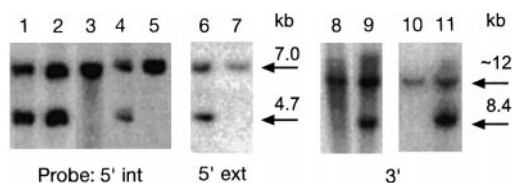


Figure 4 Southern analysis of nuclear transfer lambs. Lanes 1–5 show *Bam*HI-digested genomic DNA samples hybridized with 5' internal and external probes as indicated. Arrows indicate the 7-kb *Bam*HI fragment from the non-targeted ovine *COL1A1* locus, and the 4.7-kb fragment from the COLT-1 and COLT-2 targeted locus. Lanes 6–11 show *KpnI*, *AspI* double-digested genomic DNA samples hybridized with the 3' probe. Arrows indicate the ~12-kb *AspI* fragment from the non-targeted locus and the 8.4-kb *KpnI*, *AspI* fragment from the COLT-1 and COLT-2 targeted locus. Lane 1, nuclear transfer (n.t.) lamb from cell clone PDCOL6; lane 2, n.t. lamb from cell clone PDCAAT90; lane 3, normal lamb; lane 4, cell clone PDCAAT90; lanes 5, 7 and 8, non-targeted PDFF2 cells; lane 6, n.t. lamb from cell clone PDCOL6; lane 9, cell clone PDCAAT90; lane 10, normal lamb; lane 11, n.t. lamb from cell clone PDCOL6.

Ovine genomic DNA was prepared from cell pellets, tail samples of live lambs and umbilical cord samples of dead lambs. Southern analysis was carried out by standard procedures. Three hybridization probes were used: a 5' internal probe, a 3-kb *COL1A1* SalI, *SspI* fragment corresponding to the 5' homologous arm of COLT-1 and 2 (Fig. 1); a 5' external probe, a 520-bp ovine *COL1A1* fragment directly adjacent to but outside the 5' homologous arm; a 3' internal probe, a 0.7-kb *COL1A1* SalI, *PstI* fragment immediately 3' of the integration site (Fig. 1). The diagnostic fragments detected were: a 4.7-kb *Bam*HI fragment extending across the 5' junction of the targeted locus from a *Bam*HI site within the IRES-*neo* region to a *Bam*HI site in the *COL1A1* gene 5' of the region contained in the vector; a 8.4-kb *Kpn*I, *Asp*I fragment extending across the 3' junction of the targeted locus from a *Kpn*I site within the vector to an *Asp*I site in the *COL1A1* gene flank 3' of the region contained in the vector.

Hormonal induction of lactation

Milk samples were obtained from immature ewes by hormonal induction of lactation, essentially as described²⁵.

Received 28 January; accepted 5 May 2000.

- Thompson, S., Clarke, A. R., Pow, A. M., Hooper, M. L. & Melton, D. W. Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell* **56**, 313–321 (1989).
- Campbell, K. H. S., McWhir, J., Ritchie, W. A. & Wilmut, I. Sheep cloned by nuclear transfer from a cultured cell line. *Nature* **389**, 64–66 (1996).
- Schnieke, A. S. *et al.* Human Factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts. *Science* **278**, 2130–2133 (1997).
- Cibelli, J. B. *et al.* Cloned transgenic calves produced from nonquiescent fetal fibroblasts. *Science* **280**, 1256–1258 (1998).
- Lin, F. L., Sperle, K. & Sternberg, N. Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences. *Proc. Natl Acad. Sci. USA* **82**, 1391–1395 (1985).
- Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. & Kucherlapati, R. S. Insertion of DNA sequences into the human chromosomal β -globin locus by homologous recombination. *Nature* **317**, 230–234 (1985).
- Yancz, R. J. & Porter, A. C. Therapeutic gene targeting. *Gene Ther.* **5**, 149–159 (1998).
- Prockop, D. J. & Kivirikko, K. I. Collagens: molecular biology, diseases, and potentials for therapy. *Annu. Rev. Biochem.* **64**, 403–434 (1995).
- Stacey, A. *et al.* Perinatal lethal osteogenesis imperfecta in transgenic mice bearing an engineered mutant pro- α 1(I) collagen gene. *Nature* **332**, 131–136 (1988).
- Jang, S. K. *et al.* A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.* **62**, 2636–2643 (1988).
- Carver, A. S. *et al.* Transgenic livestock as bioreactors: stable expression of human α -1-antitrypsin by a flock of sheep. *BioTechnology* **11**, 1263–1270 (1993).
- Sedivy, J. M. & Dutraux, A. Gene targeting and somatic cell genetics: a rebirth or coming of age? *Trends Genet.* **15**, 88–90 (1999).
- Mooslehner, K. & Harbers, K. Two mRNAs of mouse pro α 1(I) collagen differ in the size of the 3'-untranslated region. *Nucleic Acids Res.* **16**, 773 (1988).
- Chu, M. L., de Wet, W., Bernard, M. & Ramirez, F. Fine structural analysis of the human pro- α 1(I) collagen gene. Promoter structure, AluI repeats, and polymorphic transcripts. *J. Biol. Chem.* **260**, 2315–2320 (1985).
- Rippe, R. A., Umezawa, A., Kimball, J. P., Breindl, M. & Brenner, D. A. Binding of upstream stimulatory factor to an E-box in the 3'-flanking region stimulates α 1(I) collagen gene transcription. *J. Biol. Chem.* **272**, 1753–1760 (1998).
- Määttä, A., Ekholm, E. & Penttinen, R. P. Effect of the 3'-untranslated region on the expression levels and mRNA stability of α 1(I) collagen gene. *Biochim. Biophys. Acta* **1260**, 294–300 (1995).
- Renard, J. P. *et al.* Lymphoid hypoplasia and somatic cloning. *Lancet* **353**, 1489–1491 (1999).
- Hill, J. R. *et al.* Clinical and pathologic features of cloned transgenic calves and fetuses (13 case studies). *Theriogenology* **51**, 1451–1465 (1999).
- Kono, T. Influence of epigenetic changes during oocyte growth on nuclear reprogramming after nuclear transfer. *Reprod. Fertil. Dev.* **10**, 593–598 (1998).
- McClennaghan, M. *et al.* Production of human α -1-antitrypsin in the milk of transgenic sheep and mice targeting expression of cDNA sequences to the mammary gland. *Anim. Biotechnol.* **2**, 161–176 (1991).
- Warburton, M. J., Kimbell, R., Rudland, P. S., Ferns, S. A. & Barraclough, R. Control of type IV collagen production in rat mammary epithelial and myoepithelial-like cells. *J. Cell. Physiol.* **128**, 76–84 (1986).
- Kay, M. A. & High, K. Gene therapy for the hemophilias. *Proc. Natl Acad. Sci. USA* **96**, 9973–9975 (1999).
- Lai, L.-W. & Lien, Y.-H. Homologous recombination based gene therapy. *Exp. Nephrol.* **7**, 11–14 (1999).
- McWhir, J. *et al.* Selective ablation of differentiated cells permits isolation of embryonic stem cell lines from murine embryos with a non-permissive genetic background. *Nat. Genet.* **14**, 223–226 (1996).
- Ebert, K. D. *et al.* Induction of human tissue plasminogen activator in the mammary gland of transgenic goats. *BioTechnology* **12**, 699–702 (1994).

Acknowledgements

We would like to acknowledge the contributions of Y. Gibson and K. Mycock for embryo manipulation; E. Emslie and L. Hutchison for molecular biology technical assistance; T. Johnston for protein analysis; and the PPL Therapeutics large animal team for animal husbandry and veterinary procedures. We thank I. Garner and D. Ayares for useful discussions.

Correspondence and requests for materials should be addressed to A.J.K. (e-mail: akind@ppl-therapeutics.com).

Gigantism in mice lacking suppressor of cytokine signalling-2

Donald Metcalf, Christopher J. Greenhalgh, Elizabeth Viney, Tracy A. Willson, Robyn Starr, Nicos A. Nicola, Douglas J. Hilton & Warren S. Alexander

The Walter and Eliza Hall Institute of Medical Research and The Cooperative Research Centre for Cellular Growth Factors, Post Office, Royal Melbourne Hospital, Victoria 3050, Australia

Suppressor of cytokine signalling-2 (SOCS-2) is a member of the suppressor of cytokine signalling family, a group of related proteins implicated in the negative regulation of cytokine action through inhibition of the Janus kinase (JAK) signal transducers and activators of transcription (STAT) signal-transduction pathway¹. Here we use mice unable to express SOCS-2 to examine its function *in vivo*. SOCS-2^{-/-} mice grew significantly larger than their wild-type littermates. Increased body weight became evident after weaning and was associated with significantly increased long bone lengths and the proportionate enlargement of most organs. Characteristics of deregulated growth hormone and insulin-like growth factor-I (IGF-I) signalling, including decreased production of major urinary protein, increased local IGF-I production, and collagen accumulation in the dermis, were observed in SOCS-2-deficient mice, indicating that SOCS-2 may have an essential negative regulatory role in the growth hormone/IGF-I pathway.

We isolated genomic clones corresponding to three independent loci from two murine libraries using a SOCS-2 coding region complementary DNA as hybridization probe. Comparison of sequence from these clones with that of the SOCS-2 cDNA revealed that one locus, which consisted of three exons and two introns, encoded the predicted SOCS-2 RNA (Fig. 1a). The two other loci

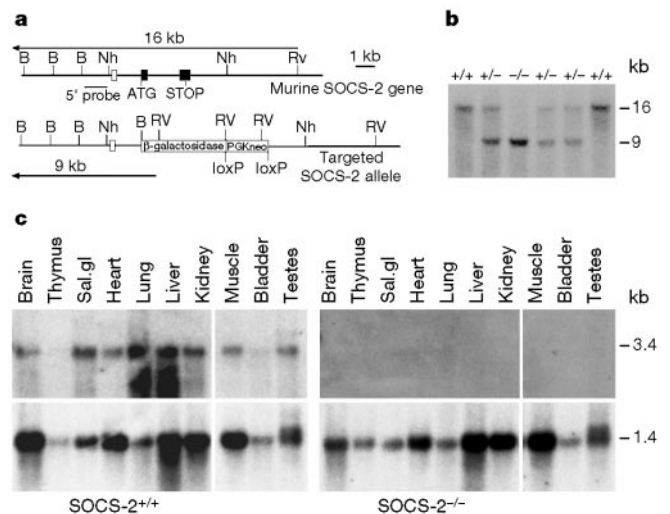


Figure 1 Disruption of the SOCS-2 locus by homologous recombination. **a**, The functional murine SOCS-2 gene (B, *Bam*HI; Nh, *Nhe*I; RV, *Eco*RV) with the exons containing the coding region as shaded boxes. In the targeted allele, the entire SOCS-2 coding region was replaced by a β -gal-PGKneo cassette in which the β -galactosidase coding region was fused to the SOCS-2 initiation codon. **b**, Southern blot of *Eco*RV-digested genomic DNA from the tails of mice derived from a cross between SOCS-2^{+/-} mice. The blot was hybridized with the 5' genomic SOCS-2 probe, which distinguishes between endogenous (16 kb) and mutant SOCS-2 (9 kb) alleles. **c**, Northern blot showing lack of SOCS-2 expression in organs of SOCS-2^{-/-} mice. Top, the blot was hybridized with a coding region probe, which detects the 3.4-kb SOCS-2 transcript¹; bottom, the integrity of the RNA was confirmed by hybridization with GAPDH (1.4-kb transcript). Sal gl, salivary gland.

represented processed or rearranged genes. Only one of these additional loci retained an intact SOCS-2 open reading frame, but it lacked the upstream region, exon 1 and part of exon 2. No cDNA clones corresponding to RNAs expressed from these rearranged loci were detected. We constructed a targeting vector for the deletion of the two coding exons in the SOCS-2 gene by homologous recombination in embryonic stem cells (Fig. 1a). Chimaeric mice were generated from an embryonic stem cell clone bearing the targeted locus, and animals lacking one functional SOCS-2 gene were bred. Southern blot analysis at weaning revealed that offspring of heterozygous parents included mice of each of the three expected genotypes (Fig. 1b) in approximately mendelian proportions (22:51:22 for SOCS-2^{+/+}:SOCS-2^{+/-}:SOCS-2^{-/-}). Northern blots of RNA extracted from a range of organs confirmed that SOCS-2 transcripts were absent in homozygous mutant mice (Fig. 1c). This result is consistent with a lack of expression from the alternative SOCS-2 loci and confirmed that the SOCS-2 gene had been functionally deleted.

SOCS-2-deficient mice were indistinguishable from their littermates until weaning at three weeks of age but subsequently grew more rapidly (Figs 2a, 3a). SOCS-2^{-/-} males weighed significantly more than their wild-type counterparts at six weeks of age and as adults were, on average, 40% heavier. Adult SOCS-2^{+/-} males exhibited an intermediate body weight (-/-: 36.6 ± 1.0 g; +/-:

30.8 ± 1.3 g; +/-: 27.1 ± 1.6 g, at 12 weeks of age, *n* = 5–20 mice per group). Increased growth was also significant but less marked in female SOCS-2^{-/-} mice (Fig. 2a). Adult SOCS-2^{-/-} females typically attained the weight of wild-type male mice, but heterozygous SOCS-2 females were not significantly heavier than sex-matched wild-type littermates (-/-: 26.2 ± 1.5 g; +/-: 21.8 ± 0.7 g; +/-: 20.5 ± 1.4 g, at 12 weeks of age, *n* = 7–20 mice per group).

Visual examination and measurement of abdominal fat mass indicated that neither male nor female SOCS-2^{-/-} mice accumulated significantly more fatty tissue than wild-type mice. Rather, increased body weight in these mice resulted from an increase in the weight of most visceral organs (Fig. 2b) of similar magnitude to the increase in overall body weight. Sexually dimorphic or male-specific organs were not disproportionately enlarged. Carcass weight was also increased, indicating that muscle and bone may contribute significantly to the increased size of SOCS-2-deficient mice. Consistent with this interpretation, the femur, tibia, radius and humerus in SOCS-2^{-/-} mice were all significantly longer than in wild-type controls (Table 1). Body length in male SOCS-2-deficient mice was also greater, although tail length was normal (Table 1). The mean frequency of hepatic nuclei per 10 high-power fields in SOCS-2^{-/-} liver sections was no greater than that in wild-type mice (27.3 ± 4.4, *n* = 4 versus 27.7 ± 2.6, *n* = 4), and striated muscle cell width was normal in the thighs of SOCS-2-deficient animals. Thus,

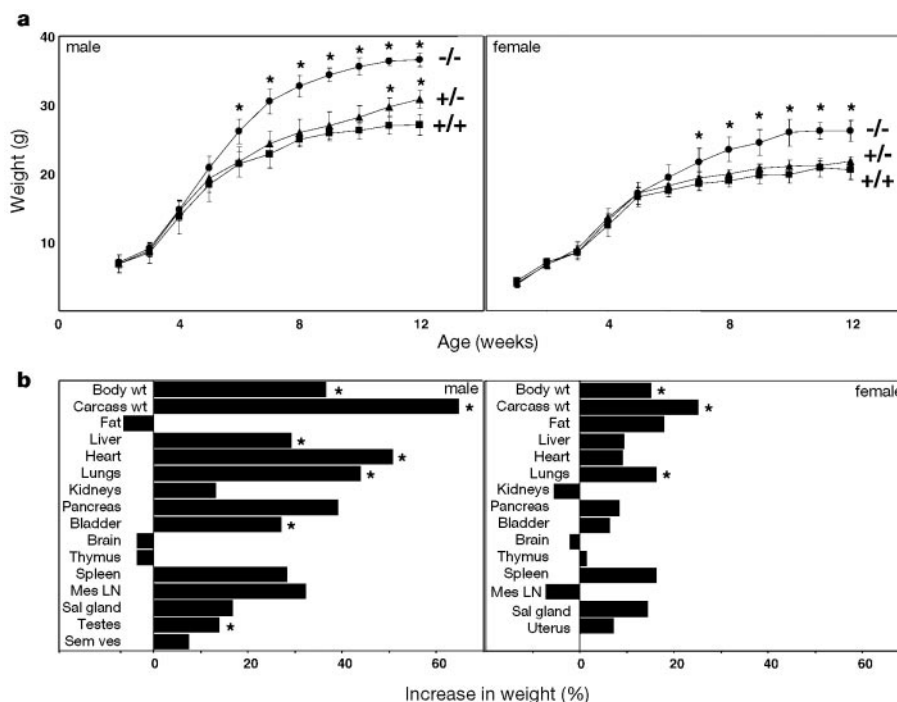


Figure 2 Excessive growth of SOCS-2^{-/-} mice. **a**, Growth curves for male and female SOCS-2^{+/+} (squares), SOCS-2^{+/-} (triangles) and SOCS-2^{-/-} (circles) mice. Body weights from cohorts of mice weighed at weekly intervals are shown: each point represents mean ± s.d. for 5–20 mice. **b**, Percentage increase in body, carcass and organ weights in

SOCS-2^{-/-} mice over that determined in age-matched wild-type mice (*n* = 7–8 three-month-old mice per measurement). Asterisk, measurements from SOCS-2^{-/-} mice that were significantly different from wild-type controls (*P* < 0.05, Student's *t*-test). wt, weight; mes LN, mesenteric lymph node; sal gland, salivary gland; sem ves, seminal vesicles.

Table 1 Body, tail and bone lengths in SOCS-2^{-/-} mice

	Male			Female		
	SOCS-2 ^{+/+}	SOCS-2 ^{+/-}	SOCS-2 ^{-/-}	SOCS-2 ^{+/+}	SOCS-2 ^{+/-}	SOCS-2 ^{-/-}
Body	95.0 ± 1.9	98.8 ± 1.6*	107 ± 0.9*	90.1 ± 2.1	91.3 ± 1.2	100.3 ± 2.1*
Tail	88.3 ± 1.5	88.7 ± 2.1	89.2 ± 1.5	84.8 ± 1.4	88.1 ± 1.6	88.9 ± 1.9*
Femur	18.0 ± 0.2	18.7 ± 0.2*	19.5 ± 0.2*	17.3 ± 0.2	17.7 ± 0.2*	18.6 ± 0.3*
Tibia	20.0 ± 0.2	20.6 ± 0.2*	20.9 ± 0.1*	19.5 ± 0.2	19.9 ± 0.2*	20.5 ± 0.3*
Radius	12.0 ± 0.2	12.4 ± 0.2*	12.7 ± 0.2*	11.6 ± 0.2	11.7 ± 0.2	12.0 ± 0.2*
Humerus	13.8 ± 0.2	14.4 ± 0.3*	15.4 ± 0.2*	13.2 ± 0.2	13.7 ± 0.3*	14.6 ± 0.1*

* *P* < 0.005 in Student's *t*-test for comparison of sex-matched SOCS-2^{-/-} and SOCS-2^{+/-} data with SOCS-2^{+/+} mice (*n* = 6–12).

increased organ weights in these mice apparently result from elevated cell numbers rather than increased cell size. A similar but less pronounced trend was also observed when organ and carcass weights and body, tail and bone lengths were assessed in female SOCS-2^{-/-} mice (Fig. 2b and Table 1).

In adult SOCS-2^{-/-} mice, most organs appeared histologically normal, including the kidney, heart, spleen, thymus, lymph nodes, femur, sternum, gonads and bladder. We did, however, observe a marked thickening of the dermis associated with excessive collagen accumulation in all of seven male, and less prominently in six of eight female, SOCS-2^{-/-} mice (Fig. 3b, c). Collagen deposition was also increased around the bronchi and vessels of the lungs of six of nine male, but only three of ten female, SOCS-2-deficient mice; this occasionally also involved some alveolar sacs (Fig. 3d, e). In a minority of SOCS-2-deficient animals of both sexes, excess collagen accumulation was also observed around occasional hepatic vessels and bile ducts, as well as in duct tissue of the salivary glands and pancreas. No abnormalities in bone architecture were evident in histological sections of adult mice, including the epiphyseal growth plates of the femur and tibia.

SOCS-2 expression can be induced by stimulation of haemopoietic tissues with cytokines¹. However, a survey of 14 SOCS-2^{-/-} mice at 2 months of age did not identify any haematological abnormalities. The haematocrit and numbers of platelets, lymphocytes, monocytes, neutrophils and eosinophils in the peripheral blood of SOCS-2^{-/-} mice were normal. The spleen weight and cellularity of the femoral bone marrow and the peritoneal cavity were increased in SOCS-2-deficient mice, but only in parallel with their elevated body weight, and the percentages of morphologically identifiable cells in cytocentrifuge preparations of cells from each of these sites were normal. Flow cytometric analysis of cells from bone

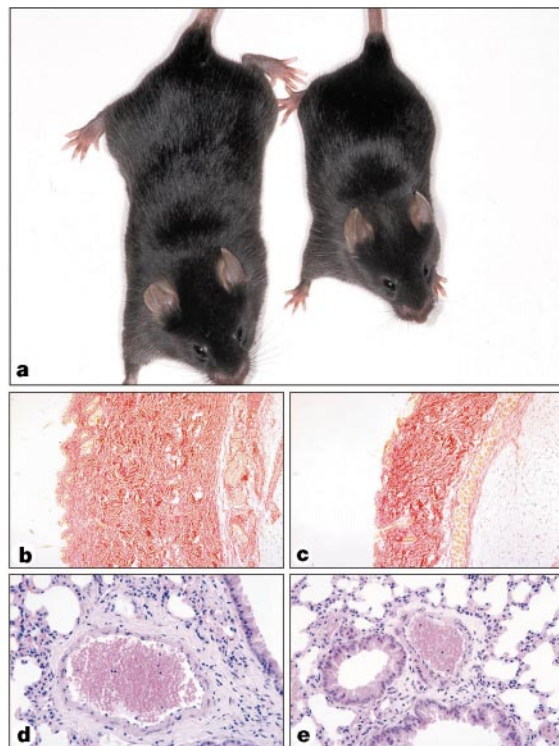


Figure 3 Pathology in SOCS-2^{-/-} mice. **a**, Increase in size of a typical two-month-old male SOCS-2^{-/-} mouse (left) relative to an age- and sex-matched wild-type animal (right). **b**, Van Gieson-stained section of skin from a two-month-old male SOCS-2^{-/-} mouse showing dermis thickened by increased collagen deposits. **c**, Skin of age- and sex-matched wild-type mouse. **d**, Haematoxylin and eosin-stained section showing collagen deposition around a lung vessel in a two month-old male SOCS-2^{-/-} mouse. **e**, Lung of age- and sex-matched wild-type mouse.

marrow, spleen and thymus, using monoclonal antibodies specific for a range of T- and B-lymphoid, myeloid and erythroid markers, revealed no consistent perturbations in SOCS-2-deficient mice. Similarly, cultures of bone marrow and spleen cells from five SOCS-2^{-/-} and five wild-type mice showed similar total colony numbers and proportions of colony subtypes when independently stimulated by GM-CSF (granulocyte-macrophage colony-stimulation factor), G-CSF (granulocyte-CSF), M-CSF (macrophage-CSF), interleukin (IL)-3, stem cell factor (SCF), IL-6 or Flk-ligand plus leukaemia inhibitory factor (LIF) (data not shown).

As members of the SOCS family can inhibit signals emanating from cytokine receptors², an attractive hypothesis for the excess growth of SOCS-2^{-/-} mice is that SOCS-2 is a negative regulator of signalling from growth-promoting cytokines. Growth hormone is a key regulator of post-natal growth, acting largely through production of IGF-I (refs 3, 4). The accelerated growth of SOCS-2^{-/-} mice, beginning at about 3–4 weeks of age, coincides with the upregulation of tissue growth hormone receptor expression^{5,6}, and the excess bone growth seen in SOCS-2^{-/-} mice is a specific characteristic of growth hormone transgenic mice that is also observed in humans with elevated growth hormone^{7,8}. However, the adult weights and rates of growth in SOCS-2^{-/-} mice more closely resemble those seen in IGF-I transgenic mice⁷. The excess collagen deposition observed in the dermis of SOCS-2^{-/-} mice is a common characteristic of both growth hormone and IGF-I transgenic mice^{9,10}, and of humans with

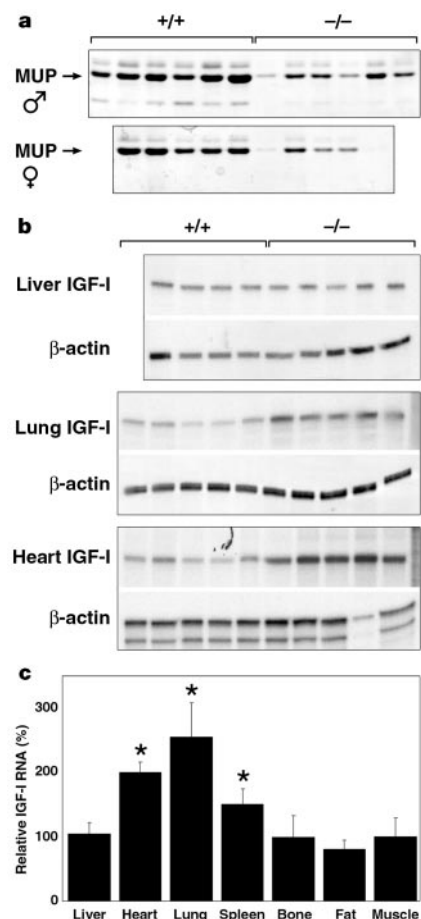


Figure 4 Deregulated growth hormone signalling in SOCS-2^{-/-} mice. **a**, Decreased MUP in urine samples from male and female SOCS-2^{-/-} mice. **b**, RNase protection assays of IGF-I expression in tissues of SOCS-2^{-/-} mice. In **a** and **b**, each lane represents a sample from an individual SOCS-2^{-/-} or age- and sex-matched wild-type mouse. **c**, Expression of IGF-I RNA in tissues of male SOCS-2^{-/-} mice, expressed as a percentage of expression in age- and sex-matched wild type mice. Data represents mean $\pm$ s.d. for comparison of at least four pairs of mice. Asterisk, $P < 0.05$.

excess growth hormone⁸.

These observations indicated that aspects of growth hormone and/or IGF-I signalling might be deregulated in SOCS-2^{-/-} mice. To examine the growth hormone/IGF-I pathway, we first investigated levels of major urinary protein (MUP), a growth hormone pulse-dependent product that is downregulated in transgenic mice over-expressing growth hormone, where the usual pulsatile pattern of growth hormone signalling is disrupted¹¹. Consistent with deregulated growth hormone signalling, there was less MUP in samples of urine from each of six male and five female SOCS-2^{-/-} mice than in a similar number of sex-matched wild-type samples (Fig. 4a). Production of IGF-I is also stimulated by growth hormone, and, consistent with deregulated growth hormone action, RNase protection assays revealed increased IGF-I production in several organs including the heart, lungs and spleen of SOCS-2^{-/-} mice (Fig. 4b, c). Excess production was not, however, evident in the liver, bone, fat or muscle. We observed no increase in serum IGF-I concentration in SOCS-2^{-/-} mice (male -/-: 278 ± 74 ng ml⁻¹; +/-: 282 ± 45; female -/-: 310 ± 62; +/-: 298 ± 76; n = 6 mice per group), consistent with normal production in the liver, which is the major source of circulating IGF-I (refs 12, 13).

The excess growth in SOCS-2^{-/-} mice reveals a key physiological role for SOCS-2 in the control of postnatal growth by growth hormone/IGF-I. SOCS-2 is induced by growth hormone *in vitro* and *in vivo* and, at least in some tissue culture assays, has a concentration-dependent inhibitory effect on growth hormone signalling¹⁴⁻¹⁶. Our data provide strong evidence that SOCS-2 is an essential negative regulator of growth hormone signalling *in vivo*, and that its absence leads to increased growth, at least in some organs, through increased local production of IGF-I. The observation that IGF-I production is normal in the livers of SOCS-2^{-/-} mice supports previous studies in which specific deletion of IGF-I in the liver led to low serum concentrations without affecting growth^{12,13}. This reinforces the emerging model that autocrine/paracrine actions of IGF-I are of paramount importance. The absence of elevated IGF-I production in some organs of SOCS-2^{-/-} mice may be explained by tissue-specific differences in the induction of SOCS genes by growth hormone^{16,17}. As other SOCS genes can be induced by growth hormone and inhibit its actions, at least when overexpressed^{15,18,19}, other members of the SOCS family may compensate for the loss of SOCS-2 regulation of growth hormone in certain tissues, resulting in normal IGF-I transcription²⁰. Nevertheless, as most organs in SOCS-2^{-/-} mice were enlarged, including some in which elevated IGF-I was not detected, our data may indicate that SOCS-2 is also indispensable in regulating IGF-I signalling itself. Indeed, SOCS-2 has been shown to interact with the IGF-I receptor^{21,22}. A role for SOCS-2 in regulating both growth hormone and IGF-I signalling in organ-specific contexts is consistent with our observation that SOCS-2^{-/-} mice exhibit characteristics of both growth hormone and IGF-I transgenic mice without entirely recapitulating either phenotype. More detailed analyses of growth hormone and IGF-I signalling in SOCS-2-deficient mice, as well as the precise definition of the temporal and spatial patterns of SOCS-2 expression in response to these cytokines, will further clarify the role of SOCS-2 in growth control.

Methods

Generation of targeted embryonic stem cells and mutant mice

We used polymerase chain reaction (PCR) to generate a genomic SOCS-2 fragment extending ~2.0 kilobases (kb) from the protein initiation ATG. This fragment was fused to the ATG of β-galactosidase using the *Bam*HI site in the plasmid vector pβgalpAloxneo²³. The 3' arm, an *Eco*RI fragment extending 3.7 kb downstream from the termination codon, was blunted and ligated into the *Xho*I (blunted) site of pβgalpAloxneo that already contained the 5' arm. This targeting vector was linearized with *Nor*I and electroporated into C57BL/6 embryonic stem cells. Transfected cells were selected in G418 and resistant clones picked and expanded. We identified clones in which the targeting vector had recombined with the endogenous SOCS-2 gene by hybridizing *Eco*RV-digested genomic DNA with a 0.8-kb *Bam*HI-*Nhe*I fragment situated in the 5' SOCS-2 genomic sequence

just outside the targeting vector (Fig. 1). This probe distinguished between the endogenous (16 kb) and targeted (9 kb) SOCS-2 alleles. A targeted embryonic stem cell clone was injected into Balb/c blastocysts to generate chimaeric mice. Male chimaeras were mated with C57BL/6 females to yield SOCS-2 heterozygotes, which were interbred to produce wild-type (SOCS-2^{+/+}), heterozygous (SOCS-2^{+/-}) and mutant (SOCS-2^{-/-}) mice on a pure C57BL/6 genetic background. We determined the genotypes of offspring by Southern blot analysis of genomic DNA extracted from tail biopsies. The deletion of the SOCS-2 coding sequence and subsequent inability to produce SOCS-2 messenger RNA in mutant mice was confirmed in nucleic acid blots performed as described²⁴. Northern blots were probed with a full-length SOCS-2 coding region probe and then with a 1.2-kb *Pst*I chicken glyceraldehyde-3-phosphate dehydrogenase (GAPDH) fragment.

Haematological and histological analyses

Peripheral blood white cell and platelet counts were determined manually using haemocytometers. Single-cell suspensions from femoral bone marrow and spleen were prepared, and differential counts of peripheral blood, bone marrow and spleen were performed from stained smears and cytocentrifuge preparations. Clonal cultures of 2.5 × 10⁴ adult bone marrow cells were performed in 0.3% agar as described²⁵. Cultures were stimulated with recombinant purified GM-CSF, G-CSF, M-CSF, IL-3 (each at 10 ng ml⁻¹), SCF (100 ng ml⁻¹), IL-6 (100 ng ml⁻¹) or Flk-ligand (500 ng ml⁻¹) plus LIF (10³ U ml⁻¹). Agar cultures were fixed and sequentially stained for acetylcholinesterase, Luxol fast blue and haematoxylin, and the cellular composition of each colony was determined microscopically. Tissue sections were prepared by standard techniques, stained with haematoxylin and eosin, and examined by light microscopy.

MUP analysis

To analyse MUP levels, 6–7-week-old mice were made to urinate before samples were collected 3 and 5.5 h later. Samples were pooled and centrifuged (13,000g × 3 min) before 0.5 μl of supernatant was electrophoresed in 12% SDS-polyacrylamide gels and stained with Coomassie blue.

Growth curves and linear measurements

We weighed cohorts of mice at weekly intervals for 12 weeks from birth. After being killed, animals were pinned down through the oral cavity and lightly stretched by the tail for nose–anus (body length) and anus–tail (tail length) measurements. To measure skeletal dimensions, limbs were subsequently removed and oriented in a consistent manner for X-ray photography and bone length measurement.

IGF-I measurements

We determined IGF-I levels in serum from orbital bleeds using an EIA kit (rat IGF DSL-10-2900, Diagnostic Systems Laboratories) according to the manufacturer's instructions. IGF-I RNA expression was determined in RNase protection assays as described¹³ using β-actin as an internal standard.

Received 7 April; accepted 11 May 2000

1. Starr, R. *et al.* A family of cytokine-inducible inhibitors of signalling. *Nature* **387**, 917–921 (1997).
2. Hilton, D. J. Negative regulators of cytokine signal transduction. *Cell. Mol. Life Sci.* **55**, 1568–1577 (1999).
3. Schoenle, E., Zapf, J., Humbel, R. E. & Froesch, E. R. Insulin-like growth factor I stimulates growth in hypophysectomized rats. *Nature* **296**, 252–253 (1982).
4. Jones, J. I. & Clemmensen, D. R. Insulin-like growth factors and their binding proteins: biological actions. *Endocrine Rev.* **16**, 3–34 (1995).
5. Shobe, L., An, M. R., Frank, S. J. & Lowe, W. L. Developmental regulation of insulin-like growth factor-I and growth hormone receptor gene expression. *Mol. Cell. Endocrinol.* **152**, 125–136 (1999).
6. Mathews, L. S., Enberg, B. & Norstedt, G. Regulation of rat growth hormone receptor gene expression. *J. Biol. Chem.* **264**, 9905–9910 (1989).
7. Kopchick, J. J., Bellush, L. L. & Coschigano, K. T. Transgenic models of growth hormone action. *Annu. Rev. Nutrition* **19**, 437–461 (1999).
8. Daughaday, W. H. in *Endocrinology* (ed. DeGroot, L. J.) 303–329 (W. B. Saunders Company, Philadelphia, 1995).
9. Quaife, C. J. *et al.* Histopathology associated with elevated levels of growth hormone and insulin-like growth factor I in transgenic mice. *Endocrinology* **124**, 40–48 (1989).
10. Wanke, R. *et al.* Overgrowth of skin in growth hormone transgenic mice depends on the presence of male gonads. *J. Invest. Dermatol.* **113**, 967–971 (1999).
11. Norstedt, G. & Palminter, R. Secretory rhythm of growth hormone regulates sexual differentiation of mouse liver. *Cell* **36**, 805–812 (1984).
12. Yakar, S. *et al.* Normal growth and development in the absence of hepatic insulin-like growth factor I. *Proc. Natl Acad. Sci. USA* **96**, 7324–7329 (1999).
13. Sjogren, K. *et al.* Liver-derived insulin-like growth factor I (IGF-I) is the principal source of IGF-I in blood but is not required for postnatal body growth in mice. *Proc. Natl Acad. Sci. USA* **96**, 7088–7092 (1999).
14. Adams, T. E. *et al.* Growth hormone preferentially induces the rapid, transient expression of SOCS-3, a novel inhibitor of cytokine receptor signaling. *J. Biol. Chem.* **273**, 1285–1287 (1998).
15. Favre, H., Benhamou, A., Finidori, J., Kelly, P. A. & Edery, M. Dual effects of suppressor of cytokine signaling (SOCS-2) on growth hormone signal transduction. *FEBS Lett.* **453**, 63–66 (1999).
16. Tollet-Egnell, P., Flores-Morales, A., Stavreus-Evers, A., Sahlin, L. & Norstedt, G. Growth hormone regulation of SOCS-2, SOCS-3, and CIS messenger ribonucleic acid expression in the rat. *Endocrinology* **140**, 3693–3704 (1999).
17. Davey, H. W., McLachlan, M. J., Wilkins, R. J., Hilton, D. J. & Adams, T. E. STAT5b mediates the growth hormone-induced expression of SOCS-2 and SOCS-3 mRNA in the liver. *Mol. Cell. Endocrinol.* **158**, 111–116 (1999).

18. Ram, P. A. & Waxman, D. J. SOCS/CIS protein inhibition of growth hormone-stimulated STAT5 signaling by multiple mechanisms. *J. Biol. Chem.* **274**, 35553–35561 (1999).
19. Hansen, J. A., Lindberg, K., Hilton, D. J., Nielsen, J. H. & Billestrup, N. Mechanism of inhibition of growth hormone receptor signaling by suppressor of cytokine signaling proteins. *Mol. Endocrinol.* **13**, 1832–1843 (1999).
20. Matsumoto, A. *et al.* Suppression of STAT5 functions in liver, mammary glands, and T cells in cytokine-inducible SH2-containing protein 1 transgenic mice. *Mol. Cell. Biol.* **19**, 6396–6407 (1999).
21. Dey, B. R., Spence, S. L., Nissley, P. & Furlanetto, R. W. Interaction of human suppressor of cytokine signaling (SOCS)-2 with the insulin-like growth factor-I receptor. *J. Biol. Chem.* **273**, 24095–24101 (1998).
22. Lopaczynski, W. Differential regulation of signaling pathways for insulin and insulin-like growth factor I. *Acta Biochim. Pol.* **46**, 51–60 (1999).
23. Starr, R. *et al.* Liver degeneration and lymphoid deficiencies in mice lacking suppressor of cytokine signalling-1. *Proc. Natl Acad. Sci. USA* **95**, 14395–14399 (1998).
24. Alexander, W. S., Metcalf, D. & Dunn, A. R. Point mutations within a dimer interface homology domain of c-mpl induce constitutive receptor activity and tumorigenesis. *EMBO J.* **14**, 5569–5578 (1995).
25. Alexander, W. S., Roberts, A. W., Nicola, N. A., Li, R. & Metcalf, D. Deficiencies in progenitor cells of multiple hematopoietic lineages and defective megakaryocytopoiesis in mice lacking the thrombopoietic receptor c-mpl. *Blood* **87**, 2162–2170 (1996).

Acknowledgements

We thank N. Sprigg, J. Mighall, S. Mifsud and L. DiRago for technical assistance; K. Hanzinikolas for animal care; S. Mihajlovic for histology; and F. Koentgen, L. Barnett, J. DeWinter and M. Swift of the Walter and Eliza Hall Institute Genetically Modified Mouse Facility and T. Adams for discussions and critical assessment of the manuscript. This work was supported by the National Health and Medical Research Council, Canberra, the Anti-Cancer Council of Victoria, an Australian Government Cooperative Research Centres Program Grant, the National Institutes of Health, Bethesda, the J. D. and L. Harris Trust and AMRAD Operations Pty Ltd, Melbourne.

Correspondence and requests for materials should be addressed to W.S.A. (e-mail: alexander_w@wehi.edu.au).

Crystal structure of enteropathogenic *Escherichia coli* intimin–receptor complex

Yu Luo[†], Elizabeth A. Frey^{*,†}, Richard A. Pfuetzner^{*,†}, A. Louise Creagh[†], Derek G. Knoechel^{*,†}, Charles A. Haynes[†], B. Brett Finlay^{*,†} & Natalie C. J. Strynadka^{*,†}

^{*} Department of Biochemistry and Molecular Biology, and [†] Biotechnology Laboratory, University of British Columbia, Vancouver V6T 1Z3, British Columbia, Canada

Intimin and its translocated intimin receptor (Tir) are bacterial proteins that mediate adhesion between mammalian cells and attaching and effacing (A/E) pathogens. Enteropathogenic *Escherichia coli* (EPEC) causes significant paediatric morbidity and mortality world-wide¹. A related A/E pathogen, enterohaemorrhagic *E. coli* (EHEC; O157:H7) is one of the most important food-borne pathogens in North America, Europe and Japan. A unique and essential feature of A/E bacterial pathogens is the formation of actin-rich pedestals beneath the intimately adherent bacteria and localized destruction of the intestinal brush border². The bacterial outer membrane adhesin, intimin³, is necessary for the production of the A/E lesion and diarrhoea⁴. The A/E bacteria translocate their own receptor for intimin, Tir⁵, into the membrane of mammalian cells using the type III secretion system. The translocated Tir triggers additional host signalling events and actin nucleation, which are essential for lesion formation. Here we describe the crystal structures of an EPEC intimin carboxy-terminal fragment alone and in complex with the EPEC Tir intimin-binding domain, giving insight into the molecular mechanisms of adhesion of A/E pathogens.

Intimin has an amino-terminal bacterial membrane anchor and

C-terminal domains needed for Tir binding⁶ (Fig. 1). Tir spans the host cell membrane with both its N and C termini in the host cytoplasm and an extracellular intimin-binding domain (IBD)⁷. The crystal structure of the C-terminal EPEC intimin fragment (residues 658–939) at 1.9 Å resolution has three adjacent domains (Fig. 2): immunoglobulin-like (Ig) D1 (residues 658–751) and D2 (residues 752–841), and a C-type lectin-like D3 (residues 842–939). The secondary structural elements and domain boundaries are comparable to the previously determined NMR structure⁸, but the overall rod shape of intimin contrasts sharply with the curved NMR architecture (Fig. 2). The observed rod shape of intimin is also maintained in the crystallographic complex with the Tir IBD.

Intimin makes different and minimal crystal packing contacts in the two crystal forms (the complex crystal has a high solvent content of 73%). Thus, the observed elongated shape is unlikely to be a crystallographic artefact. In addition, the crystal structure of intimin matches closely the 2.3 Å resolution crystal structure of the extracellular fragment of *Yersinia pseudotuberculosis* invasin (residues 503–986), a homologue of intimin that binds to β_1 integrins. Invasin has four Ig-like domains (D1–D4) and a C-type lectin-like domain (D5)⁹. Despite a low sequence identity of only 21%, 241 out of the 282 C α atoms of our intimin structure were matched to invasin with a moderate root mean square (r.m.s.) deviation of 2.9 Å using ALIGN¹⁰. Intimin D2 and D3 form a superdomain with 1,572 Å² of buried surface (calculated using GRASP¹¹), much like invasin D4 and D5. Like the Ig-like domain interfaces of invasin, the intimin D1–D2 interface accounts for 450 Å² of buried surface. The intimin D1 aligns equally well (~1.3 Å r.m.s. deviation for ~90% matched C α atoms) with invasin D1, D2 and D3, implying a similar architecture with multiple Ig-like repeats. Furthermore, the N-terminal 539 amino acids of intimin and the N-terminal 489 amino acids of invasin (33% identity) are interchangeable and sufficient to promote bacterial surface localization of the C-terminal fragments⁶, implying that both these adhesins have a similar outer

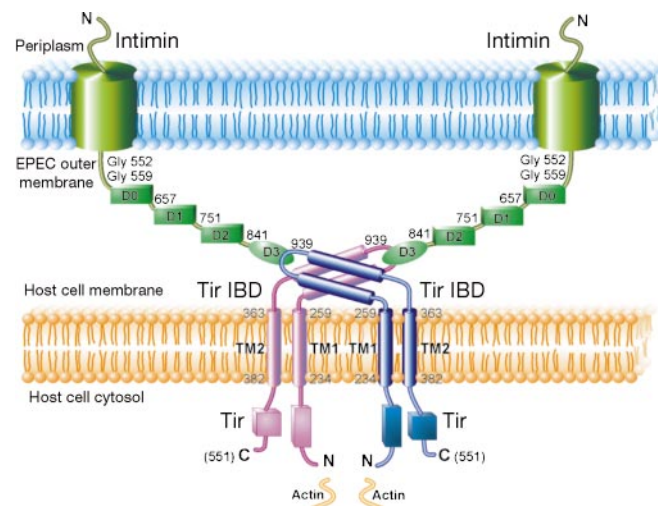


Figure 1 The EPEC/host-cell adhesion interface. The model is based on our structural data of the complex of the C-terminal fragment of intimin (domains D1, D2 and D3) and the extracellular Tir IBD. Intimin is shown in green with domains labelled and boundary residues numbered. The Ig-like domains D0, D1 and D2 are shown as rectangles, and the lectin-like domain D3, which binds to the Tir IBD, as an oval. Tir is shown as a dimer (in pink and dark blue) in the host-cell membrane, and is also labelled and numbered as described for intimin. The Tir IBD is the extracellular component of Tir flanked by the two predicted transmembrane (TM) domains. We observe a dimeric Tir IBD, with the two helices in each monomer forming a four-helix bundle that is stabilized by multiple hydrophobic and hydrogen-bonded interactions. The N-terminal domain of Tir anchors host cytoskeletal components (such as actin) that are needed to form the characteristic A/E lesion on the host-cell surface upon bacterial adhesion.

membrane anchor. Using the dense alignment surface method¹², we predict the intimin residues 542–550 and the invasins residues 492–500 are the last transmembrane segments in both adhesins. The intimin segment (residues 559–657) shares 22–24% sequence identity with each of the first three Ig-like domains of invasins (as well as 35% sequence identity with the intimin Ig-like domain D1), and the hinge region linking this segment to D1 has a sequence pattern similar to its invasin D1–D2 counterpart.

On the basis of these analogies, we predict that the intimin extracellular segment has one additional Ig-like domain, which we term D0 (residues 559–657), which probably extends the structure of our experimentally observed 282-residue fragment in a similar fashion to that observed with the rod-shaped invasin (Fig. 2). As such, the entire extracellular segment of intimin would be an elongated and relatively rigid rod with approximate dimensions of $140 \text{ \AA} \times 30 \text{ \AA} \times 30 \text{ \AA}$, about $40 \text{ \AA}$ shorter than invasin (which has an additional Ig-like domain). Thus it appears that both adhesins use tandem Ig-like domains to create an appropriate length of rod, and a C-terminal lectin-like domain to interact with their receptors. Finally, the first Ig domain of invasin, D1, is linked to its last putative transmembrane β -strand by only a two-residue linker (Pro 501 and Gln 502). Intimin, however, has a longer eight-residue segment with a conserved glycine residue near each end (Gly 552 and Gly 559), which would link the first Ig-domain, D0, to the last residue of the putative transmembrane anchor. We speculate that the conformational variability of the glycine residues may allow this region to serve as a flexible hinge mediating mechanical movement between the rigid extracellular arm (domains D0–D3) and the bacterial outer membrane to which it is anchored (Fig. 1).

The structure of the EPEC Tir IBD (residues 272–362) in complex with intimin was determined at $2.9 \text{ \AA}$ resolution (Fig. 2). Residues following Ala 336 were disordered in the map, matching limited proteolysis studies which showed that this region was the most labile in the IBD–intimin complex, with a chymotrypsin cleavage site after Val 337 (data not shown). We observe an intimately associated crystallographic Tir IBD dimer (Fig. 2). Each Tir IBD monomer is hairpin-shaped with two long α -helices (HA, residues 274–291 and HB, residues 315–334), a three-residue 3_{10} helix initiated by Pro 294, and a β -hairpin (residues 297–311). The two α -helices mediate dimerization by forming an antiparallel four-helix bundle similar to ColE1 Rop protein¹³ (r.m.s. deviation on 80 common C α atoms is $3.2 \text{ \AA}$). A surface area of $1,971 \text{ \AA}^2$ is buried at the dimerization interface (Figs 2, 3); this is significantly more than usually found at a dimer, which is simply an artefact of crystallographic packing¹⁴. The interface has a hydrophobic core (except for a pair of hydrogen-bonded Thr 282 side chains) and a rim of

polar and charged side chains (57% of the buried surface area is contributed by nitrogen and oxygen atoms with nine direct hydrogen bonds between them). These features match the characteristics of native helical bundle proteins¹⁵, indicating that there may be a biological role for Tir dimerization. EPEC Tir has two predicted transmembrane α -helices (Fig. 1; TM1: 234–259, TM2: 363–382)¹². The proposed extracellular central domain between the two transmembrane α -helices is the most highly conserved segment amongst all ten available Tir sequences (69% identity between EPEC and EHEC Tir IBD), emphasizing the stringent structural requirements for intimin binding and the observed dimerization interface. TM1 and HA are linked by 12 residues with an unusual and highly conserved sequence (residues 260–271): TPE(A)PDD(S)PT(I)TTDP. Considering that there is one such linker sequence at each end of the antiparallel four-helix bundle, we predict that this proline-rich segment could serve as a rigid anchor of a dimeric IBD to TM1. Mutagenesis, crosslinking, hydrodynamic and fluorescence resonance energy transfer (FRET) studies can now be used to address Tir dimerization *in vitro* and *in vivo*.

Isothermal titration calorimetry results indicate that EPEC Tir IBD and EPEC intimin have a binding constant, K_a , of $3.2 \times 10^6 \text{ M}^{-1}$ at 37°C , comparable to that of full-length Tir and intimin (data not shown). Our structure reveals that binding of intimin and Tir is mediated primarily by interactions between the lectin-like D3 domain of intimin and the β -hairpin and the N terminus of helix

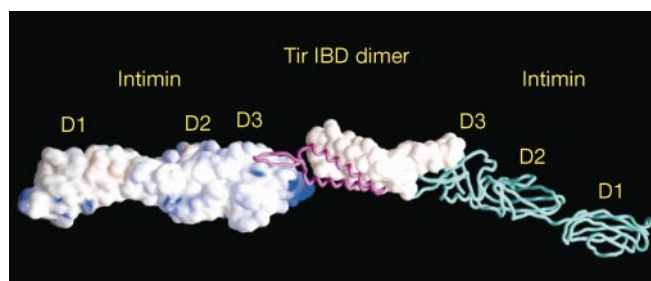


Figure 3 GRASP¹¹ surface representation of the dimeric intimin–Tir IBD complex. The viewing direction is approximately parallel to the dimerization dyad. Accessible surfaces colour-coded with electrostatic potential (-15 for red, $+10$ for blue) are shown for one intimin (on the left) and one Tir IBD (in the centre). The other intimin (in blue) and Tir IBD (in pink) are shown as worm models. Although Tir IBD has an overall net negative charge (seven net negative charges) the dimerization interface between the two Tir molecules is minimally charged. Intimin has a complementary overall positive charge (six net positive charges) with a positively charged tip close to the α -helices of the Tir IBD dimer.

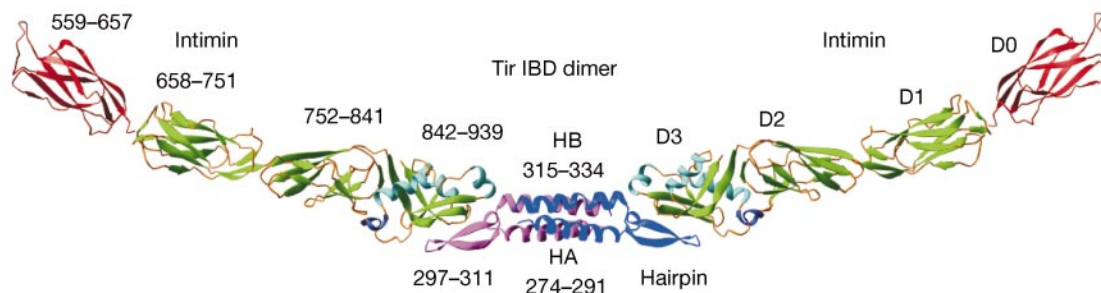


Figure 2 A RIBBONS³⁰ representation of the intimin and Tir IBD dimer complex. The D1, D2 and D3 domains of the two intimin monomers are colour-coded by secondary structures (α -helices in cyan, β -strands in green, 3_{10} -helices in purple, coils in brown). The predicted Ig-like domain D0, whose orientation was modelled based on the invasin structure⁹, is in red. The Tir IBD dimer in the centre is colour-coded by chains (pink and dark blue, respectively). The viewing direction is perpendicular to the vertical dyad axis. All 282 C α atoms of the high resolution structure of free intimin could be superposed on the

Tir IBD-bound form of intimin with an r.m.s. deviation of $0.67 \text{ \AA}$. The 280 C α atoms of the previously determined NMR structure of intimin⁸ were superposed on the crystal structure with an r.m.s. deviation of $16.3 \text{ \AA}$, due primarily to differences in domain orientation. We could align 155 of the 282 C α atoms (all in D2 and D3) with the NMR structure with a r.m.s. deviation of $3.75 \text{ \AA}$. D1, D2 and D3, which have 78, 81 and 90 matched C α atoms, respectively, gave r.m.s. deviations of $3.09 \text{ \AA}$, $2.58 \text{ \AA}$ and $3.53 \text{ \AA}$.

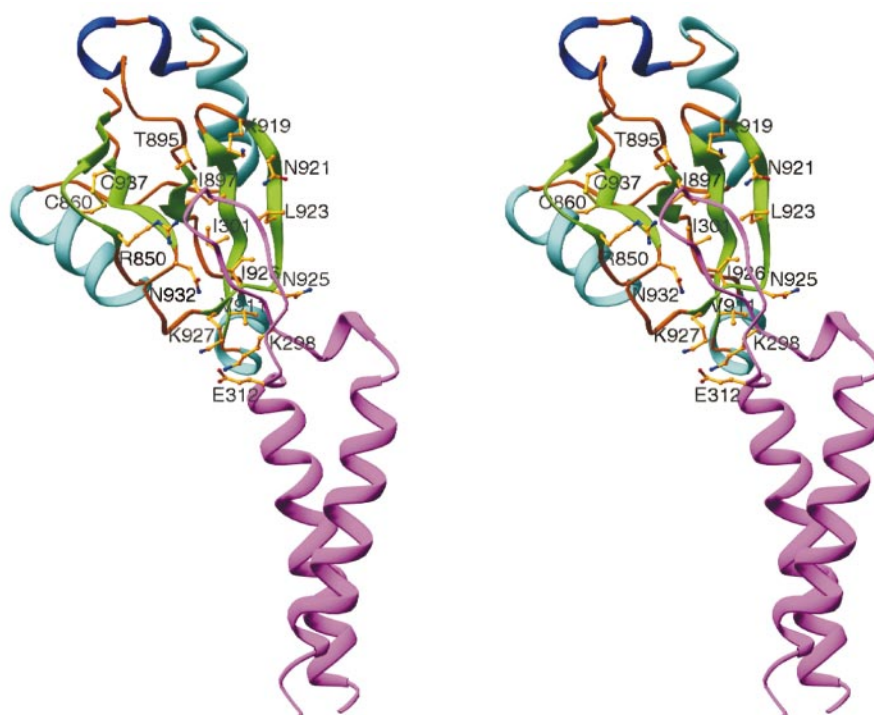


Figure 4 Stereo RIBBONS³⁰ representation of intimin D3 (colour-coded by secondary structures) and Tir IBD (in pink) with side chains of the residues involved in intimin–Tir

recognition. The α -hairpin of Tir IBD is shown as coil. The previously identified disulphide bridge (Cys 860–Cys 937)⁸ distant from the binding interface is also shown.

HB of Tir IBD (Figs 3 and 4). The 1,335 Å² of buried interface is generally flat and uncharged, except for the tip of intimin which has an overall positive charge. The extensive hydrophobic nature of the large interface probably explains the favourable enthalpy change ($\Delta H = -40.1 \text{ kcal mol}^{-1}$) and the large heat capacity change ($C_p = -1.7 \text{ kcal mol}^{-1} \text{ K}^{-1}$) that occurs upon binding (data not shown). The positively charged intimin tip is formed by a short α -helix (residues 904–909) which is hydrogen-bonded through the main-chain carbonyl oxygens of Ser 909 and Lys 908 to the main-chain amides of Asp 315 and Asp 316 within HB of Tir.

A direct salt bridge is formed between intimin Lys 927 and Tir Glu 312 at the surface of the interface. Both of these charged residues are in close proximity to Tir Lys 298 (Fig. 4). It is interesting that in

the related pathogen EHEC, Lys 298 is conserved within Tir whereas the residue at position 312 is a non-polar valine. In addition, EHEC intimin has an asparagine at position 927. In a cross-strain interaction, the electrostatic repulsion between Lys 298 of EHEC Tir and Lys 927 of EPEC intimin in the absence of the complementary Glu 312 helps to explain EHEC Tir's 20-fold reduced affinity towards EPEC intimin compared with its natural partner¹⁶. The side chain of EPEC intimin Val 911 is entirely buried in the binding interface (Fig. 4). Mutation of this residue would disrupt packing at the interface, explaining the observed deleterious effect on Tir-binding of the EHEC intimin V906A mutant (equivalent to V911 of EPEC intimin)⁶. Beneath Tir's β -hairpin, the unusual main-chain conformation of Gln 920 enables the side chains of Lys 919, Leu 923,

Table 1 Crystallographic data

Data collection	Native	SeMet	Me ₃ PbAc	HgAc ₂	Intimin/IBD
Data set:					
Resolution (Å)	1.9 (1.94–1.9)	2.2 (2.28–2.2)	2.4 (2.49–2.4)	2.4 (2.49–2.4)	2.9 (2.96–2.9)
Reflections	24,955 (875)	17,935 (1,742)	13,966 (1,395)	13,945 (1,402)	16,811 (821)
Redundancy	3.73	3.57	6.66	5.97	5.08
Completeness	90.7 (51.5)	99.6 (96.8)	99.9 (98.7)	99.9 (99.9)	96.3 (74.9)
<i>I</i> / σ <i>I</i>	15.1 (2.7)	21.6 (5.7)	15.1 (3.6)	12.5 (3.1)	12.2 (2.0)
<i>R</i> _{merge}	0.097 (0.261)	0.069 (0.207)	0.105 (0.366)	0.078 (0.300)	0.095 (0.283)
Phasing statistics					
Sites		2 Se	1 Pb	1 Hg	
<i>R</i> _{cryst} , iso/ano		0.84/–	0.81/0.98	0.59/0.82	
Phasing power, iso/ano		0.93/–	1.01/0.55	2.83/0.99	
Refinement statistics					
Data set	Intimin	Intimin–Tir IBD			
Model composition	282 res + 190 water	348 res, no water			
Reflections in working/test sets	22,103/2,440	14,272/1,578			
<i>R</i> -factor/free <i>R</i> -factor:	0.215/0.255	0.247/0.296			
Bond (Å)/angle (°) deviation	0.0051/1.419	0.0078/1.391			
Ramachandran plot					
Most favoured regions	225/89.3%	248/79.7%			
Additional allowed regions	25/9.9%	57/18.3%			
Generously allowed regions	1/0.4%	5/1.6%			
Disallowed regions	1/0.4%	1/0.3%			

Data for high-resolution bins are listed in parentheses. Phasing statistics calculated by SHARP²³ were calculated to 2.4 Å. Regions on the Ramachandran plot are defined by PROCHECK²⁹.

Asn 925 and Ile 926 to form a flat interface with the side chains of Arg 850, Thr 895, Ile 897 and Asn 932. Intimin Ile 926 and Tir Ile 301 are buried in the centre of the interface (Fig. 4). The extensive interactions observed at the Tir–intimin interface may indicate a fixed orientation of the IBD β -hairpin.

Both intimin and invasin have a C-type lectin-like C-terminal domain. In relation to the putative integrin-binding site of invasin⁹, the Tir binding site of intimin is located at the opposite side of the C-terminal lectin-like domain. Invasin lacks the short α -helix (residues 904–909 of intimin) involved in Tir binding. On the other hand, intimin lacks the charge constellation thought to be required for binding integrin⁹. Thus, our work shows that despite the overall sequence and structural similarities in the rod-forming, tandem Ig-like domains and the overall fold of the C-terminal recognition domain, these two related adhesins have evolved distinctive sites for recognizing their respective receptors. EPEC intimin has a pair of cysteine residues (860 and 937, Fig. 4) that are conserved in both invasin and all C-type lectins. A disulphide bridge between these residues is observed in our structure of the intimin–IBD complex but not in the structure of free intimin owing to different rotameric states of Cys 937. This disulphide may not be necessary to maintain a C-type lectin fold (although it may stabilize it to some degree). The two cysteine residues are both distant from the Tir-binding site. We therefore speculate that the EPEC intimin C937S mutant⁷ would probably preserve the major structural features required for binding Tir but may induce sufficient destabilization of the protein fold to affect the overall affinity of the interaction *in vivo*. Finally, EPEC intimin lacks the usual cluster of calcium-binding acidic residues that has been shown to be a critical feature of the carbohydrate binding site in the C-type lectins¹⁷. Thus, if intimin does bind to host cell-surface carbohydrate, it must do so in a significantly different manner. In the absence of any other supporting data, we feel that the observation of a lectin-like fold is insufficient to support the notion of intimin binding to host cell-surface carbohydrate, as has been suggested⁷.

We have observed an antiparallel Tir IBD–intimin dimer with explicit rigidity. The structural model points towards a novel mode of adhesin–receptor binding with both proteins horizontally orientated in the intermembrane space, anchoring A/E bacteria intimately to the membrane of the host cell (Fig. 1). This mode of binding provides an alternative explanation for the inability to label intimin at the attachment interface with polyclonal antibodies *in vivo*¹⁸. The largely nonpolar surface of intimin with an overall positively charged tip would allow it to sit favourably on top of the membrane or anionic sialylated cellular mask. The potential significance of the observed Tir four-helix bundle dimerization domain, the primarily hydrophobic intimin–Tir binding interface and intimin and its transmembrane segment can now be addressed by mutagenesis. On the basis of sequence analysis, intimin's N-terminal 550-residue putative membrane anchor region may form a porin-like structure¹⁹, as predicted for its homologue invasin⁹. The N-terminal domain of Tir has been mapped to bind the host cytoskeletal component α -actinin²⁰. We speculate that a multivalent complex formed by Tir dimerization (as suggested by the crystal structure) and intimin trimerization (as observed in several bacterial outer membrane porins¹⁹) could mediate the formation of symmetrically packed actin fibrils that aid in creation of the observed lesion in the host cell beneath adherent A/E bacteria. □

Methods

Sample preparation

We amplified the EPEC intimin C-terminal fragment 658–939 from chromosomal DNA by polymerase chain reaction (PCR) and inserted it into vector pET-21a. The transformed *E. coli* strain BL21 (DE3) was grown at 37 °C to OD 0.4 and induced with 0.5 mM IPTG overnight at 25 °C. The seleno-methionine derivative of intimin was expressed with *E. coli* strain 834 (DE3) in 2 × M9 medium with 40 mg l⁻¹ seleno-methionine and 19 non-methionine

amino-acids. We passed the sonicated cell lysate in 20 mM cacodylate-HCl buffer at pH 6.7 through a DE52 anion-exchange medium to trap anionic proteins and nucleic acids. The protein in the flow-through was purified by a Mono-S cation-exchange column using 0–200 mM NaCl gradient. EPEC Tir fragment 272–362 (IBD) was PCR-amplified and inserted into vector pET-28a with an N-terminal hexahistidine affinity tag. It was purified by nickel-chelating Sepharose, subjected to thrombin cleavage to remove the affinity tag and further purified by a Mono-Q anion-exchange column using a 0–500 mM NaCl gradient. Purified intimin and IBD were co-concentrated in 10 mM Tris-HCl buffer with 30 mM NaCl at pH 8.2. We removed excess unbound intimin or IBD by size-exclusion chromatography with a Superdex-200 column. Purified proteins at ~5 mg ml⁻¹ were not concentrated further.

Crystallization

We used hanging-drop vapour diffusion to screen crystallization conditions. Intimin alone crystallized in 1.8 M K₂HPO₄ or 2.5 M (NH₄)₂SO₄ at pH ~7.5. Crystals grew to a suitable size of 0.4 mm × 0.3 mm × 0.03 mm in sitting drops about three days after seeding with crushed crystals. Heavy atom derivatives of HgAc₂ and (Me)₃PbAc were prepared by soaking. The intimin–IBD complex crystallized readily in either 1.8 M (NH₄)₂SO₄ with 200 mM Tris-HCl buffer at pH 8.5 or 24% PEG 8000 and 100 mM MgSO₄ with the same buffer. Crystals grew to a maximum size of 0.5 mm × 0.2 mm × 0.05 mm in about three weeks.

Data collection

All data were collected at 100 K in an Oxford cryogenic stream. The cryo-protectants had 25% glycerol plus the crystallization recipe. We recorded oscillation images using a Rigaku Raxis-II dual image plate system mounted on a RU-200 X-ray generator equipped with osmic mirrors. Data were processed using DENZO²¹. The isomorphous intimin crystals crystallized in K₂HPO₄ or (NH₄)₂SO₄ belong to space group P2₁ with cell dimensions $a = 52.16$ Å, $b = 47.10$ Å, $c = 71.66$ Å and $\beta = 94.2^\circ$. The intimin–IBD crystals in PEG 8000 and (NH₄)₂SO₄ were also isomorphous. They belong to space group C222₁ with cell dimensions $a = 90.5$ Å, $b = 349.0$ Å, $c = 47.8$ Å. However, the crystal in PEG 8000 did not diffract to better than 4.5 Å. Statistics are listed in Table 1.

Structure determination

Initial heavy atom positions were located using CCP4 programs²². Multiple isomorphous and anomalous scattering (MIRAS) phase probabilities with a combined figure of merit of 0.6 to 2.6 Å resolution were obtained using SHARP²³. The subsequent phase modified map (SOLOMON²⁴) was easily traced with XFIT²⁵ and O²⁶. We refined the intimin structure (one molecule per asymmetric unit, 57% solvent) with CNS²⁷. Gln 920, the only Ramachandran outlier, has clear density. The intimin–IBD structure was solved using AMoRe²⁸ with refined intimin as the search model. The 3.0 Å molecular replacement model-phased map revealed 66 residues of uninterrupted density with most sidechains. The complex model (one intimin and one Tir IBD per asymmetric unit with 73% solvent) was also refined with CNS. The refinement statistics are listed in Table 1.

Received 1 March; accepted 10 May 2000.

- Todd, E. C. Epidemiology of foodborne diseases: a worldwide review. *World Health Stat. Q.* **50**, 30–50 (1997).
- Knutton, S. Cellular responses to enteropathogenic *Escherichia coli* infection. *Biosci. Rep.* **15**, 469–479 (1995).
- Jerse, A. E., Yu, J., Tall, B. D. & Kaper, J. B. A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. *Proc. Natl Acad. Sci. USA* **87**, 7839–7843 (1990).
- Donnenberg, M. S. *et al.* Role of the *eaeA* gene in experimental enteropathogenic *Escherichia coli* infection. *J. Clin. Invest.* **92**, 1412–1417 (1993).
- Kenny, B. *et al.* Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell* **91**, 511–520 (1997).
- Liu, H., Magoun, L., Luperchio, S., Schauer, D. B. & Leong, J. M. The Tir-binding region of enterohaemorrhagic *Escherichia coli* intimin is sufficient to trigger actin condensation after bacterial-induced host cell signalling. *Mol. Microbiol.* **34**, 67–81 (1999).
- Hartland, E. L. *et al.* Binding of intimin from enteropathogenic *Escherichia coli* to Tir and to host cells. *Mol. Microbiol.* **32**, 151–158 (1999).
- Kelly, G. *et al.* Structure of the cell-adhesion fragment of intimin from enteropathogenic *Escherichia coli*. *Nature Struct. Biol.* **6**, 313–318 (1999).
- Hamburger, Z. A., Brown, M. S., Isberg, R. R. & Bjorkman, P. J. Crystal structure of invasin: a bacterial integrin-binding protein. *Science* **286**, 291–295 (1999).
- Satow, Y., Cohen, G. H., Padlan, E. A. & Davies, D. R. Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 Å. *J. Mol. Biol.* **190**, 593–604 (1986).
- Honig, B. & Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **268**, 1144–1149 (1995).
- Cserzo, M., Wallin, E., Simon, I., von Heijne, G. & Elofsson, A. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Prot. Eng.* **6**, 673–676 (1997).
- Banner, D. W., Kokkinidis, M. & Tsernoglou, D. Structure of the ColE1 rop protein at 1.7 Å resolution. *J. Mol. Biol.* **196**, 657–675 (1987).
- Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl Acad. Sci. USA* **93**, 13–20 (1996).
- Betz, S. F., Liebman, P. A. & DeGrado, W. F. De novo design of native proteins: characterization of proteins intended to fold into antiparallel, rod-like, four-helix bundles. *Biochemistry* **36**, 2450–2458 (1997).
- DeVinney, R. *et al.* Enterohaemorrhagic *Escherichia coli* O157:H7 produces Tir, which is translocated to the host cell membrane but is not tyrosine phosphorylated. *Infect. Immun.* **67**, 2389–2398 (1999).

17. Weis, W. I., Dickamer, K. & Hendrickson, W. A. Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature* **360**, 127–134 (1992).
18. Knutton, S. *et al.* Down regulation of intimin expression during attaching and effacing enteropathogenic *Escherichia coli* adhesion. *Infect. Immun.* **65**, 1644–1652 (1997).
19. Schirmer, T. General and specific porins from bacterial outer membranes. *J Struct Biol.* **121**, 101–109 (1998).
20. Goosney, D. L. *et al.* Enteropathogenic *E. coli* translocated intimin receptor, Tir, interacts directly with α -actinin. *Curr. Biol.* (in the press).
21. Otwinowski, Z. Oscillation data reduction program. *Proceedings of the CCP4 Study Weekend: Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 56–62 (SERC Daresbury Laboratory, Warrington, UK, 1993).
22. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
23. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
24. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
25. McRee, D. E. *Practical Protein Crystallography* (Academic, San Diego, 1993).
26. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
27. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
28. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
29. Laskowski, R. A., MacArthur, M. W. & Thornton, J. M. Validation of protein models derived from experiment. *Curr. Opin. Struct. Biol.* **8**, 631–639 (1998).
30. Carson, M. & Bugg, C. E. Algorithm for ribbon models of proteins. *J. Mol. Graphics*, **4**, 121–122 (1986).

Acknowledgements

We thank S. Mosimann for suggestions on the Se-Met intimin derivative, F. Ness for help in the production of Fig. 1, and D. Lim and S. Mosimann for software support. Work in our laboratories was supported by the Medical Research Council (MRC) of Canada, the Canadian Bacterial Disease Network Center of Excellence, Burroughs Wellcome Foundation (BWF) and the Toronto Hospital for Sick Children. N.C.J.S. is an MRC Scholar and a BWF New Investigator, and B.B.F. is an MRC Scientist and a Howard Hughes International Research Scholar.

Correspondence and requests for material should be addressed to N.C.J.S. The Protein Data Bank ID codes are 1F00 and 1F02.

Solid-state NMR determination of the secondary structure of *Samia cynthia ricini* silk

J. D. van Beek*, L. Beaulieu*, H. Schäfer†, M. Demura‡, T. Asakura§ & B. H. Meier*

* Laboratory for Physical Chemistry, ETH Zürich, CH-8092 Zürich, Switzerland

† Laboratory for Physical Chemistry, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands

‡ Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

§ Department of Biotechnology, Tokyo University of Agriculture and Technology, Koganei, Tokyo 184-8588, Japan

Silks are fibrous proteins that form heterogeneous, semi-crystalline solids. Silk proteins have a variety of physical properties reflecting their range of functions. Spider dragline silk, for example, has high tensile strength and elasticity¹, whereas other silks² are better suited to making housing, egg sacs or the capture spiral of spiders' webs. The differing physical properties arise from variation in the protein's primary and secondary structure, and their packing in the solid phase. The high mechanical performance of spider dragline silk, for example, is probably due to a β -sheet conformation of poly-alanine domains³, embedded as small crystallites within the fibre. Only limited structural information can be obtained from diffraction of silks^{3–6}, so further characterization requires spectroscopic studies

such as NMR^{7–11}. However, the classical approach to NMR structure determination¹² fails because the high molecular weight¹³, repetitive primary structure¹³ and structural heterogeneity of solid silk means that signals from individual amino-acid residues cannot be resolved. Here we adapt a recently developed solid-state NMR technique^{14,15} to determine torsion angle pairs (ϕ , ψ) in the protein backbone, and we study the distribution of conformations in silk from the Eri silkworm, *Samia cynthia ricini*. Although the most probable conformation in native fibres is an anti-parallel β -sheet, film produced from liquid directly extracted from the silk glands appears to be primarily α -helical.

The Eri silkworm *Samia cynthia ricini* produces housing silk with a primary structure that strongly resembles spider dragline silk, with repeating poly-alanine domains of (Ala)_n, where $n = 10–14$, interleaved with glycine-rich domains¹¹. Its mechanical properties, however, are closer to those of silk from *Bombyx mori* silkworm, which has a different primary structure. One-dimensional solid-state NMR studies show differences between cast films (of the liquid silk from the gland) and native fibres of silk from *S. c. ricini*, demonstrating the strong influence of the preparation on the resulting secondary structure in the solid. On the basis of

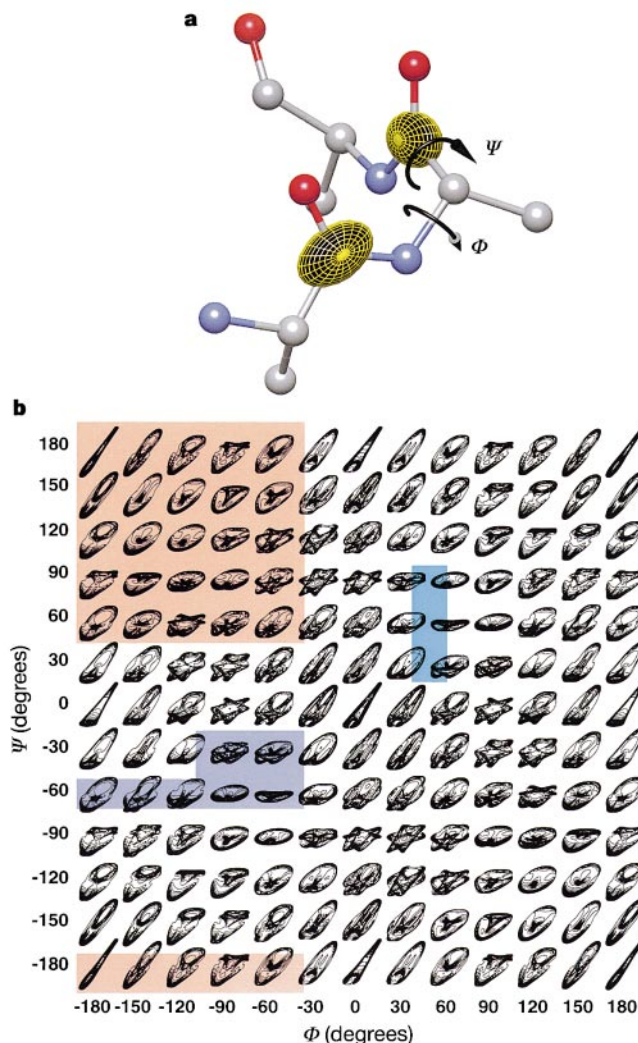


Figure 1 The (ϕ , ψ)-dependence of DOQSY spectra in the two-spin approximation. **a**, Molecular geometry of a polypeptide and orientation of the ¹³C = O CSA tensors in the molecular frame. **b**, Simulated DOQSY spectra as function of (ϕ , ψ). Energetically favourable areas (for alanine) are indicated by a coloured background: red, β -sheet; blue, α -helix; cyan, left-handed α -helix.

chemical-shift analyses, α -helical and β -sheet conformations have been postulated for the two states respectively^{10,11}. Although isotropic chemical shift values can usually distinguish between α -helical and β -sheet conformations, they are always compatible with a large number of (ϕ, ψ) combinations.

Here we use a two-dimensional (2D) NMR technique that is able to distinguish between these (ϕ, ψ) values yielding the probability density for the torsion angles, $P(\phi, \psi)$. Measurement of the relative orientation of the anisotropic chemical shielding (CSA) tensors at two nuclear positions delivers geometric information¹⁶ without requiring macroscopically orientated samples¹⁰. In the case of the protein backbone, the torsion angles (ϕ, ψ) can be determined from the relative orientation of the $^{13}\text{C}=\text{O}$ CSAs (Fig. 1), whose orientation with respect to the molecular frame is known quite precisely. Here we use a double-quantum/single-quantum correlation experiment (DOQSY)¹⁴ to correlate two carbon tensor orientations. The samples have been labelled with ^{13}C at the alanine C = O position. Nevertheless, we deal with a multi-spin system and the spin dynamics may be complex. In proteins, however, the C = O distances between carbonyl carbons in amino acids adjacent in the primary structure are, with few exceptions, shorter than other C = O

contacts. The former are found to lie between 2.7 Å and 3.7 Å depending on (ϕ, ψ) , the latter usually exceed 4.2 Å (from a search among 5,165 proteins in the Protein Data Bank). For poly-alanine in an extended anti-parallel β -sheet conformation, for example, the distances amount to 3.6 Å and 4.8 Å, respectively¹⁷. For short excitation/reconversion times (initial rate), the DOQSY excitation depends on the inverse of the sixth power of the distance, and the spectra can be approximately described by the alanine C = O pairs adjacent in the primary structure. However, signals arising from the non-adjacent contacts may be up to 25% of the total signal intensity. For the anti-parallel β -sheet, this mainly gives intensity at $(-180^\circ, 180^\circ)$ as the tensor orientations for these contacts are parallel. In initial-rate approximation, ^{13}C -carbonyls that have ^{13}C neighbours on both sides have spectra that approach those of two independent spin pairs. Furthermore, with 15% labelling as in our samples, they are about five times less probable than spin pairs. These arguments show that the initial-rate DOQSY spectra of proteins may be described by a superposition of two-spin spectra.

The initial-rate DOQSY spectrum calculated as a function of the torsion angles, (ϕ, ψ) , is shown in Fig. 1b on a grid with 30° resolution. The variation is strong, and allows for a sensitive determination of $P(\phi, \psi)$ from experimental data, except that it is not possible to distinguish left- from right-handed helices: $(\phi, \psi) \leftrightarrow (-\phi, -\psi)$. Spectra related by $(\phi, \psi) \leftrightarrow (\psi, \phi)$ are also similar because the angles between the C = O CSA tensors in the two conformations are similar. However, the marked difference in the dipolar interaction tensor orientation (with respect to the CSA) breaks the approximate symmetry (Fig. 1b), and orientations related by $(\phi, \psi) \leftrightarrow (\psi, \phi)$ are distinguishable if the signal-to-noise ratio is sufficiently high.

The experimental DOQSY spectra for both fibre and film samples and the best fits are shown in Fig. 2. The respective probability densities $P(\phi, \psi)$ for both the fibre and film are shown in Fig. 3a and 3b. Both samples have narrow, well-defined distributions. To resolve the $(\phi, \psi) \leftrightarrow (-\phi, -\psi)$ ambiguity, we used the dependence of the isotropic chemical shift of the α -carbon on (ϕ, ψ) as calculated for

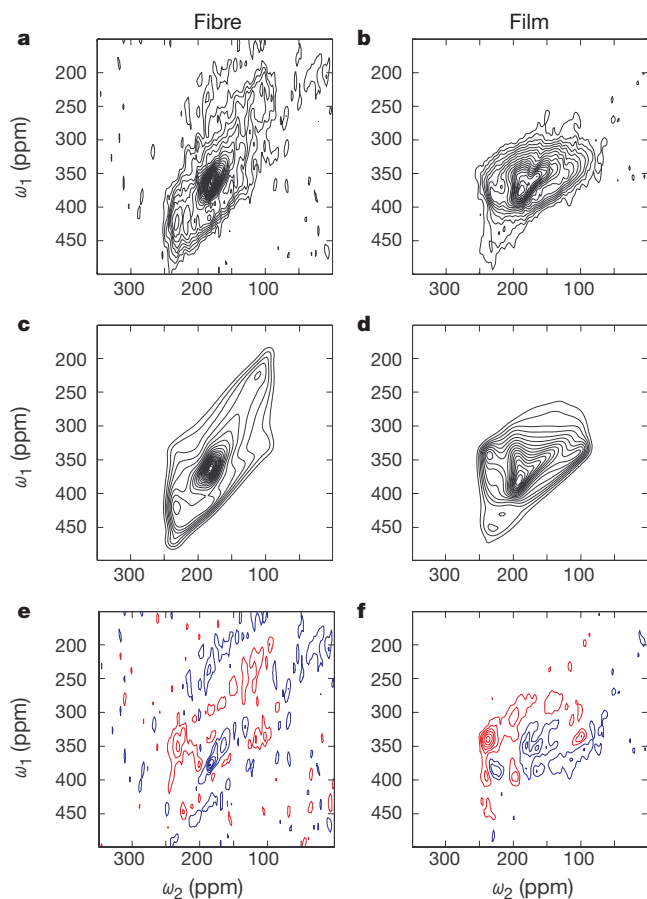


Figure 2 Experimental DOQSY spectra of silk from *Samia cynthia ricini* silkworms **a, b**, Spectra for fibre (**a**) and film (**b**) samples. **c, d**, Best fit for the fibre (**c**) and film (**d**) samples. These include the chemical-shift bias but are identical to the ones without bias at the resolution of the contour plot **e, f**. Differences between best fit and experimental data for fibre (**e**) and film (**f**). Red, negative intensity. For spectra **a–d**, the total signal intensity of the carbonyl region in the 2D spectra is normalized to 1,000 for a resolution of 1.76 ppm per point. The levels are absolute levels and are set at 0.025, 0.050, 0.075, and so on, to enable direct comparison. The principal values of the $^{13}\text{C}=\text{O}$ CSA tensors used in the fit were obtained from static 1D powder patterns and amounted to (244.8, 194.7, 91.6) ppm for the film and (244.8, 184.5, 89.2) ppm for the fibre sample.

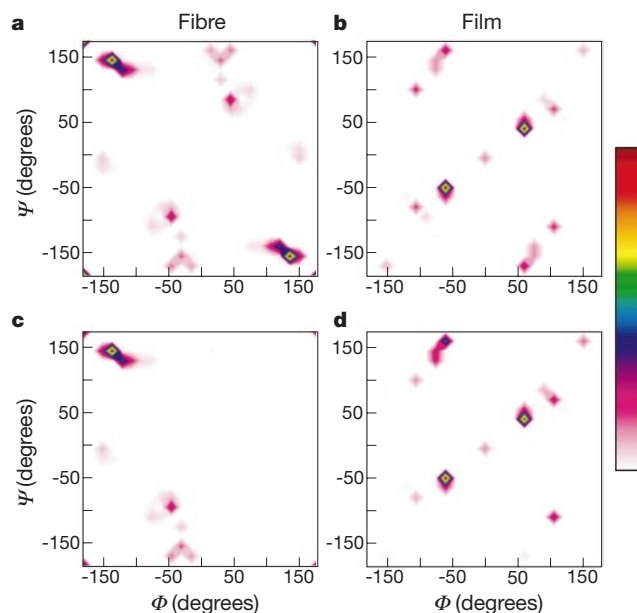


Figure 3 Probability distribution functions for (ϕ, ψ) determined from the experimental spectra in Fig. 2 using Tikhonov regularization. **a, b**, Solutions for the fibre (**a**) and film (**b**) samples. **c, d**, Best solutions for the fibre (**c**) and film (**d**) samples, including isotropic $^{13}\text{C}\alpha$ chemical shift values (53.1 ppm for the film and 49.2 ppm for the fibre). These values are in good agreement with the empirical prediction for α -helices and β -sheets (52.4 and 48.2 ppm¹¹), respectively.

alanine¹⁸ as an additional criterion. In the magic-angle spinning spectrum, both samples show a single resonance line at 49.2 and 53.1 ppm for fibre and film, respectively. Figure 3c and d shows the analyses using the chemical shift as an additional constraint. The major intensities for the fibrous sample are now inside the energetically favourable regions of the Ramachandran plot (Fig. 1). The most probable conformation is $(\phi, \psi) = (-135^\circ, 150^\circ)$, and 70% of $P(\phi, \psi)$ fall within the β -sheet region. The intensity around $(-180^\circ, 180^\circ)$ is attributed to inter-strand contacts within the β -sheet. For the film, the bimodal distribution of Fig. 3b (corresponding to left- and right-handed helices) cannot be reduced to a singly peaked distribution because the $C\alpha$ chemical-shift difference between the two conformations is only minor. We suspect, however, that the alanines are present in the right-handed α -helical form. The film shows a maximum at $(\pm 60^\circ, \pm 45^\circ)$ corresponding to 60% of $P(\phi, \psi)$. Within the 15° grid resolution of our analysis, these regions of the Ramachandran plot coincide with the standard conformation for the anti-parallel β -sheet $(-140^\circ, 135^\circ)$ and for the α -helix $(-57^\circ, -47^\circ)$. These results support earlier studies^{10,11} and extend them considerably because the DOQSY results allow determination of the angle pairs (ϕ, ψ) . We believe that the experimental scheme applied here, and variants of it, can also be used to elucidate the local structure in other solid proteins. □

Methods

Labelling

We labelled *S. c. ricini* silk fibroin biosynthetically by oral administration of an aqueous solution of 99% $1\text{-}^{13}\text{C}$ -alanine to 6- or 7-day-old silkworm larvae in the fifth instar stage. The $^{13}\text{C}=\text{O}$ enrichment of the alanine residues was about 15% in both samples (determined by 1D MAS NMR spectra). Cross labelling to glycine residues was more than five times lower than for alanine, leading to a Gly-Gly pair probability at least 25 times lower than for Ala-Ala. We did not detect cross labelling to other amino acids owing to their lower abundance. We prepared the film sample on a flat polystyrene plate with an aqueous silk fibroin solution gently extracted from the posterior silk glands of the silkworm larvae in the last stage of the fifth instar. We prepared the fibrous sample from cocoon filaments, and purified it to remove the sericin coating¹⁰.

NMR spectra

We obtained NMR spectra at room temperature on a Bruker DMX 400 spectrometer. Temperature does not affect the spectra in 2D single-quantum correlation experiments in the range from 150 to 300 K, indicating that dynamical effects are not important. We used the experimental scheme of ref. 14 with the double quantum excitation sequence from ref. 19, 20. We applied radio-frequency field strengths of 50 kHz, except for proton decoupling at 100 kHz. The cross-polarization contact time was 1.7 ms, the recycle delay 4 s. We used a single excitation cycle of 1.5 ms to generate double-quantum coherence. The total measuring time per spectrum was about seven days for $\sim 30\text{-mg}$ samples.

Determination of the probability density $P(\phi, \psi)$

We modelled the experimental DOQSY spectrum by $S(\omega_1, \omega_2) = \iint P(\phi, \psi) S(\phi, \psi, \omega_1, \omega_2) d\phi d\psi$ where $S(\phi, \psi, \omega_1, \omega_2)$ is the calculated DOQSY spectrum as a function of the torsion angles (ϕ, ψ) . We assumed standard bond lengths for peptides²¹, a planar peptide plane, and the established orientation of the carbonyl CSA tensor^{22–25} with the most shielded axis perpendicular to the peptide plane and the intermediately shielded axis 5° from the $\text{C}=\text{O}$ bond.

To evaluate $P(\phi, \psi)$ the integral equation given above has to be inverted. This is known to be an 'ill-posed' problem—for a given experimental spectrum many functions $P(\phi, \psi)$ exist which all give a fit of similar quality. To implement the 'common sense' notion that $P(\phi, \psi)$ is a smooth and positive function, we used three regularization tools: (1) discretization of $P(\phi, \psi)$; (2) non-negativity of $P(\phi, \psi)$; and (3) Tikhonov regularization²⁶. $P(\phi, \psi)$ was discretized by approximating it with the continuous function $\tilde{P}(\phi, \psi)$ which is identical to $P(\phi, \psi)$ at n^2 grid points of $\tilde{P}_{ij}$ on a 2D grid and, between grid points, is defined by a piece-wise linear interpolation scheme. Thus we can write $S(\omega_1, \omega_2) = \sum_{i,j=1}^{n^2} a_{ij}(\omega_1, \omega_2) \tilde{P}_{ij}$, where the n^2 coefficients of $\tilde{P}_{ij}$ define the probability density and $a_{ij}(\omega_1, \omega_2) = \int_{\phi_i+\Delta}^{\phi_i+\Delta+\Delta} \int_{\psi_j+\Delta}^{\psi_j+\Delta+\Delta} d\phi d\psi c_{ij}(\alpha, \beta) S(\alpha, \beta, \omega_1, \omega_2)$ are the basis spectra. Here $\Delta = \frac{360^\circ}{n}$ is the grid resolution and the coefficients define the linear interpolation between grid points in $\tilde{P}(\phi, \psi)$. A similar strategy for a 1D inverse problem is discussed in ref. 27.

We used a non-negative least squares algorithm and determined the regularization parameter λ for the Tikhonov regularization using the self-consistent method^{28,29}. We chose a grid resolution of 15° ($n = 24$). We calculated 144 sub-spectra $S(\phi, \psi, \omega_1, \omega_2)$ with 17,711 powder averaging points for each spectrum a_{ij} . We included the chemical-shift surfaces for the $C\alpha$ carbon¹⁸ in the kernel as a penalty function. We tolerated only a minimal increase in root-mean-squared error ($< 0.1\%$) and neglected the dependence of the $\text{C}=\text{O}$ CSA on (ϕ, ψ) . The uncertainty in and angle dependence of the tensor orientation are clearly smaller than the grid resolution and can be neglected. The (ϕ, ψ) dependence of the principal values on the CSA is relatively minor and strongest for the component along the $\text{C}=\text{O}$ bond where variations of about 10 p.p.m. are found. This is

smaller than the applied gaussian line-broadening of 15 ppm but it could lead to some distortion of $P(\phi, \psi)$ if structures with sizeable shift differences were contained in the same sample. We found well-localized distributions for the silk samples discussed, and this source of error is not important for the major components. The minor components in $P(\phi, \psi)$ could be distorted because the CSA values are dominated by the major components. These distortions would lead to a redistribution of $P(\phi, \psi)$ between conformations with strongly correlated DOQSY spectra. The α -helical orientation, for example, shows no high correlation with any other conformations (Fig. 1b), but the different spectra within the relatively large β -sheet region show more significant correlations.

Received 30 December 1999; accepted 4 May 2000.

- Denny, M. W. in *The Mechanical Properties of Biological Materials, Symposia of Soc. Exp. Biol.* Vol. 34 (eds Vincent, J. F. V. & Currey, J. D.) 247–272 (Cambridge Univ. Press, Cambridge, 1980).
- Vollrath, F. Spider webs and silks. *Sci. Am.* **266**, 52–58 (1992).
- Yang, Z., Grubb, D. T. & Jelinski, L. W. Small-angle X-ray scattering of spider Dragline silk. *Macromolecules* **30**, 8254–8261 (1997).
- Warwick, J. O. Comparative studies of fibroins II. The crystal structures of various fibroins. *J. Mol. Biol.* **2**, 350–362 (1960).
- Becker, M. A. et al. in *Silk Polymers—Material Science and Biotechnology* (eds Kaplan, D., Adams, W. W., Farmer, B. & Viney, C.) 185–195 (Am. Chem. Soc., Washington DC, 1994).
- Bram, A., Branden, C. I., Craig, C., Snigireva, I. & Riekel, C. X-ray diffraction from single fibers of spider silk. *J. Appl. Cryst.* **30**, 390–392 (1997).
- Saito, H. et al. High-resolution ^{13}C NMR study of silk fibroin in the solid state by cross-polarization magic-angle spinning method. Conformational characterization of silk I and silk II type forms of *Bombyx mori* fibroin by the conformational-dependent ^{13}C chemical shifts. *Macromolecules* **17**, 1405–1412 (1984).
- Simmons, A. H., Michal, C. A. & Jelinski, L. W. Molecular Orientation and two-component nature of the crystalline fraction of spider dragline silk. *Science* **271**, 84–87 (1996).
- Kümmerlen, J., van Beek, J. D., Vollrath, F. & Meier, B. H. Local structure in spider dragline silk investigated by two-dimensional spin-diffusion nuclear magnetic resonance. *Macromolecules* **29**, 2920–2928 (1996).
- Asakura, T., Ito, T., Okudaira, M. & Kameda, T. Structure of alanine and glycine residues of *Samia cynthia ricini* silk fibers studied with solid-state ^{15}N and ^{13}C NMR. *Macromolecules* **32**, 4940–4946 (1999).
- Ishida, M., Asakura, T., Yokoi, M. & Saito, H. Solvent- and mechanical-treatment-induced conformational transition of silk fibroins studied by high-resolution solid-state carbon-13 NMR spectroscopy. *Macromolecules* **23**, 88–94 (1990).
- Wüthrich, K. *NMR of Proteins and Nucleic Acids* (Wiley Interscience, New York, 1986).
- Kaplan, D. L., Adams, W. W., Viney, C. & Farmer, B. L. in *Silk Polymers—Material Science and Biotechnology* (eds Kaplan, D., Adams, W. W., Farmer, B. & Viney, C.) 2–16 (Am. Chem. Soc., Washington DC, 1994).
- Schmidt-Rohr, K. A double-quantum solid-state NMR technique for determining torsion angles in polymers. *Macromolecules* **29**, 3975–3981 (1996).
- Schmidt-Rohr, K., Hu, W. & Zumbulyadis, N. Elucidation of the chain conformation in a glassy polyester, PET, by two-dimensional NMR. *Science* **280**, 714–717 (1998).
- Ernst, M. & Meier, B. H. in *Solid State NMR of Polymers* Vol. 84 (eds Ando, J. & Asakura) 183–121 (Elsevier, Amsterdam, 1998).
- Arnott, S., Dover, S. D. & Elliott, A. Structure of β -poly-L-alanine: Refined atomic coordinates for an anti-parallel β -pleated sheet. *J. Mol. Biol.* **30**, 201–208 (1967).
- Heller, J. et al. Determination of dihedral angles in peptides through experimental and theoretical studies of alpha-carbon chemical shielding tensors. *J. Am. Chem. Soc.* **119**, 7827–7831 (1997).
- Antzutkin, O. N. & Tycko, R. High-order multiple quantum excitation in ^{13}C nuclear magnetic resonance spectroscopy of organic solids. *J. Chem. Phys.* **110**, 2749–2752 (1999).
- Yen, Y.-S. & Pines, A. Multiple-quantum NMR in solids. *J. Chem. Phys.* **78**, 3579–3582 (1983).
- Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl Acad. Sci. USA* **37**, 205–211 (1951).
- Stark, R. E., Jelinski, L. W., Ruben, D. J., Torchia, D. A. & Griffin, R. G. Carbon-13 chemical shift and carbon-13/nitrogen-15 dipolar tensors for the peptide bond: 1- ^{13}C -glycyl- ^{15}N -glycine hydrochloride hydrate. *J. Magn. Reson.* **55**, 266–273 (1983).
- Teng, Q., Iqbal, M. & Cross, T. A. Determination of the ^{13}C chemical shift and ^{14}N electric field gradient orientations with respect to the molecular frame in a polypeptide. *J. Am. Chem. Soc.* **114**, 5312–5312 (1992).
- Oas, T. G., Hartzell, C. J., McMahon, T. J., Drobny, G. P. & Dahlquist, F. W. The carbonyl carbon-13 chemical shift tensors of five peptides determined from nitrogen-15 dipole-coupled chemical shift powder patterns. *J. Am. Chem. Soc.* **109**, 5956–5962 (1987).
- Hartzell, C. J., Whitfield, M., Oas, T. G. & Drobny, G. P. Determination of the ^{15}N and ^{13}C chemical shift tensors of L-[^{13}C]Alanyl-L-[^{15}N]alanine from the dipole-coupled powder patterns. *J. Am. Chem. Soc.* **109**, 5966–5969 (1987).
- Tikhonov, A. N. & Arsenin, V. Y. *Solutions of Ill-Posed Problems*. (John Wiley, New York, 1977).
- Schäfer, H., Albrecht, U. & Richter, R. Dispersive first-order reactions I: data analysis. *Chem. Phys.* **182**, 53–60 (1994).
- Honerkamp, J. & Weese, J. Tikhonov regularization method for ill-posed problems. *Cont. Mech. Thermodyn.* **2**, 17–30 (1990).
- Weese, J. A reliable and fast method for the solution of Fredholm integral equations of the first kind based on Tikhonov regularization. *Comput. Phys. Comm.* **69**, 99–111 (1992).

Acknowledgements

This research was supported by the Swiss National Science Foundation, the European Science Foundation and the Bio-oriented Technology Research Advancement Institution, Japan (T.A.). We thank M. Ernst for experimental advice and R. Verel for discussion.

Correspondence and requests for materials should be addressed to B.H.M. (e-mail: beme@nmr.phys.chem.ethz.ch).

Catalytic converters

Enzyme assays, enzyme handling and enzyme kinetics.

Scanning microscopy

Bio-Rad www.bio-rad.com

Real-time laser scanning microscopes for kinetic applications

Based on Nikon's RCM8000 system, the RTS2000 (confocal) and RTS2000MP (multi-photon) video-rate scanning microscopes use a resonant galvanometer to achieve a scanning rate of 30–130 frames per second. This makes the machines suitable for recording calcium, pH and membrane potential in living cells on the tens of milliseconds timescale. In the RTS2000 MP, multi-photon microscopy gives good depth penetration and reduced photobleaching for extended organism lifetimes. The RTS2000 MP has been used to image rapid changes in intracellular calcium using FRAT-based cameleons and fluorescent calcium indicator fluo-3 (Fan *et al. Biophys. J.* 76, 2412–2420; 1999).

Reader Service No. 100

Chemical cart

Labconco Corporation www.labconco.com

Chemicals on the move

Labconco's Chemical Cart is fitted with two shelves with guard rails to prevent supplies and containers from sliding off. The shelves are made of phenolic board with a Melamine high-pressure laminate surface. The welded-construction frame consists of one-inch-square steel tubing and the vehicle runs on four-inch hard rubber casters. The frame is 19 inches wide, so it should fit easily through most doors and wide corridors. The Chemical Cart comes ready-to-roll with

no assembly required.

Reader Service No. 101

Laptop Cooler

Nalge Nunc www.nalgenunc.com

A cool place in the lab

An addition to the Cryoware range, the Nunc – 20 °C Laptop Cooler protects enzymes, cells and reagents maintaining temperatures of – 15 to – 20 °C for up to 4 hours on the laboratory bench. The unit will also maintain freezer temperature during the frost-free cycle, during power outages or accidental door-opening. Designed for use with new Nunc Cryogenic Vials, the new cooler provides capacity for up to twenty 1.5-ml to 2.0-ml microcentrifuge tubes or Nunc Brand Cryogenic Vials, and includes inserts that accommodate 0.2-ml to 0.5-ml microcentrifuge tubes.

Reader Service No. 102

MMP antibodies

Biogenesis www.biogenesis.co.uk

MMP antibodies for use in paraffin sections

These polyclonal antibodies are raised against matrix metalloproteinases MMP-1, 2, 3 and 9, and TIMP-1 native antigen. All are suitable for use on formalin-fixed paraffin embedded tissue following pre-treatment, and in blotting. The matrix metalloproteinases are a family of highly homologous zinc metalloproteinases which are thought to play a central role in the processing of the components of the extracellular matrix. They are implicated in physiological and pathological conditions, including rheumatoid arthritis, tissue morphogenesis, and tumour invasion and metastasis.

Reader Service No. 103

G7441 antibody

Promega www.promega.com

In situ detection of phospho-specific Akt

This Anti-pS⁴⁷³ Akt pAb detects phospho-specific (serine 473) Akt, an important enzyme in the study of cell survival. The antibody is a rabbit polyclonal designed for use in immunocytochemistry and western blot analysis. It is selective for the serine 473 phosphorylated form of Akt and does not cross-react with nonphosphorylated Akt.

Reader Service No. 104

Pulverisette 19

Fritsch www.fritsch.de

Grind out those enzymes

The Pulverisette 19 is a rapidly running cutting mill available with a variety of different

funnels, rotors, sieve cassettes and collecting vessels. A sample exhaustor with cyclone dust chamber can be used for materials that are difficult to grind. The casing can be opened by a central shutter with one hand. For easy cleaning and changing of the sieve cassette, the casing top and the cover can be opened and the cover is additionally removable. Throughput is up to 60 kg per hour, and maximum feed particle size is 70 × 80 mm.

Reader Service No. 105



A smashing machine.

Ceran hotplates

Gerhardt www.gerhardt.de

CE certified hotplates with a minimum of moving parts

Ceran, a glass ceramic often found in the modern kitchen, is used in the latest additions to the Gerhardt range of laboratory hotplates. With Ceran, heat is transferred quickly with minimal loss thanks to the material's high infrared transparency. The smooth glass ceramic surfaces of these hotplates helps in making spills less tricky to deal with, and routine cleaning is simplified. Model HCP is designed for heating liquids quickly, to a maximum of 1.8 kW with a choice of 9 heating levels. Model HCR1 incorporates an integrated stirrer and model HCR2 is equipped for heating, stirring and regulating, with one heating area for rapid heat-up and a second for maintaining heat while stirring.

Reader Service No. 106



Plate ceramics: Gerhardt's Ceran hotplate.



On a roll: Labconco's Chemical Cart.

new on the market

Pierce HAT-Check

Pierce www.piercenet.com

Keep a check on regulatory proteins

The HAT-Check (histone acetyltransferase) Assay Kit is a rapid HAT assay. HAT activity is an indicator of the cellular transcriptional activity, and the enzyme plays a major role in the control of gene expression. A number of transcriptional regulatory proteins have been found to possess intrinsic HAT activity. The HAT-Check Activity Assay is designed to be more convenient than methods involving SDS-PAGE procedures and provides quantitative and qualitative results. The assay is based on a synthetic peptide corresponding to the first 23 amino acids of histone H4 coupled through a GSGL linker to a biotin molecule as substrate.

Reader Service No. 107

Poly (ADP-ribose) polymerase

Trevigen www.trevigen.com

Take control of PARs

Trevigen's synthetic poly(ADP-ribose) polymer (PAR) is offered as a useful control for quantitating levels of PAR within cells using anti-PAR antibodies. Possible applications include assay development in apoptosis, DNA damage and repair, and drug discovery targeting the poly(ADP-ribose) polymerase enzyme. The polymer may also be used as a substrate for poly(ADP-ribose) glycohydrolase.

Reader Service No. 108

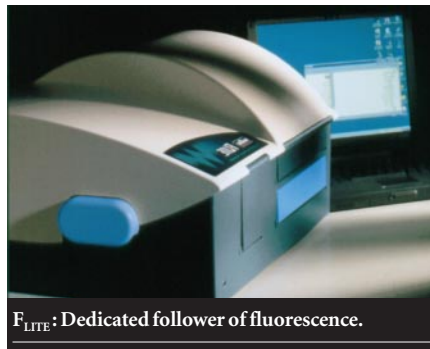
F_{LITE} fluorometer

PerkinElmer Life Sciences

www.perkinelmer.com

Fluorimetric assay

The Berthold F_{LITE} is a dedicated fluorometer featuring a range of measurement modes including kinetics, scanning, dual-ratio



F_{LITE}: Dedicated follower of fluorescence.

measurements, and top and bottom reading. Suggested applications for this instrument include cell based, binding studies, environmental toxins, quantifications, toxicology and drug screening. The F_{LITE} software allows easy measurement of up to 864-well microplates, Petri dishes, slides and filters, and Terasaki plates. F_{LITE} contains 32-bit Windows software and an RS 232 user interface.

Reader Service No. 109

FailSafe PCR System

Epicentre www.epicentre.com

PCR enzyme system for high-fidelity amplification

The FailSafe PCR System provides high-fidelity DNA amplification for both routine and difficult templates. Both long (at least 20 kb) and difficult templates (such as sequences with high GC or a high degree of secondary structure) can be amplified using a thermostable polymerase blend with 3'-to-5' proofreading activity. This system includes 12 PreMixes that contain dNTPs, buffers and various concentrations of MgCl₂ and FailSafe PCR Enhancer (with betaine). The user adds template, primers and the FailSafe PCR Enzyme Mix to the PreMixes and then amplifies to determine the optimal PreMix.

Reader Service No. 110

Equilibrium Dialyzer

AmiKa Corp. www.amika.com

For protein binding assays and screening

This 96-well Equilibrium Dialyzer is built to perform the simultaneous preparation of 96 micro-volume samples. It is designed for the study of the binding of small molecules or ions to macromolecules, such as DNA, proteins, peptides, carbohydrates and lipids. Each well has a separate membrane to reduce the risk of sample contamination and leakage. The 96-well Equilibrium Dialyzer offers precisely standardized conditions for up to 96 wells by ensuring uniformity of membrane area, sample volume, degree of agitation and temperature. The volume range for each sample is 100–200 µl per well. The system is available with membranes with MWCOs of 5,000 or 10,000 daltons.

Reader Service No. 111

Master Taq kit

Eppendorf www.eppendorf.com

Enzyme stabilization in PCR

Many PCR reactions require individual optimization of reaction conditions, as the template may form incalculable secondary structures, inhibitors may impair the reaction or other imponderables may prevent results from turning out as expected. TaqMaster is an all-purpose enzyme stabilizer which is chemically inert, thermostable and attacks neither the integrity of the template and primers nor the enzymes. Its effect on PCR assays is attributable to a higher tolerance to magnesium concentration as well as a higher tolerance of the polymerase to pH fluctuations. TaqMaster combines with Taq DNA Polymerase to form the MasterTaq kit for complex PCR problems.

Reader Service No. 112

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

Custom Peptides & Polyclonal Antibodies

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

Research Genetics
U.S. or Canada 800.531.4763
Fax 256.546.9818

Check out our homepage at
<http://www.rgen.com>

Lysyl Endopeptidase
(Achromobacter Protease 1)
Specifically cleaves all Lys-X bonds including Lys-Pro.
Highly Stable, Water Soluble

Wako

Wako BioProducts
Wako Chemicals USA, Inc.
1000 Bellwood Road
Richmond, VA 23267
Toll-free: (800) 271-7877
Fax: (804) 271-7701
(800) 952-8226
bioproducts@wako.com

Diagnostic Reagents

Petachrom[®] Series: chromogenic and fluorogenic substrates

Petabloc[®] Series: serine proteases inhibitors

Purified snake venom proteins (e.g. Protac[®], Ecarin, Batroxobin)

PENTAPHARM LTD
CH 4002 Basel Fax: +41 61 310 00 10
<http://www.pentapharm.ch>

READER ENQUIRY NO. 39

READER ENQUIRY NO. 23

READER ENQUIRY NO. 16